# (19) World Intellectual Property Organization

International Bureau



# A TREATH BRITARIA DA BRITA BRITA

(43) International Publication Date 12 August 2004 (12.08.2004)

PCT

# (10) International Publication Number WO 2004/067743 A1

(51) International Patent Classification?: 15/63, C07H 21/04, A01H 5/00

C12N 15/11,

(21) International Application Number:

PCT/US2003/041269

(22) International Filing Date:

24 December 2003 (24.12.2003)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/441,377	21 January 2003 (21.01.2003)	US
60/441,381	21 January 2003 (21.01.2003)	US
60/441,392	21 January 2003 (21.01.2003)	US
607441,405	21 January 2003 (21.01.2003)	US
60/441,447	21 January 2003 (21.01.2003)	US
60/441,502	21 January 2003 (21.01.2003)	US
10/609,019	26 June 2003 (26.06.2003)	US

- (71) Applicants (for all designated States except US): TRANS-GENRX, INC. [US/US]; 1755 Wittington Place Drive, Suite 300, Dallas, TX 75234 (US). THE BOARD OF SUPERVISORS OF LOUISIANA STATE UNIVERSITY AND AGRICULTURAL AND MECHANICAL COLLEGE [US/US]; P.O. Box 25055, Baton Rouge, LA 70894 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): COOPER, Richard,

K. [US/US]; 111 Pecan Meadow Drive, Baton Rouge, LA 70810 (US). FIORETTI, William, C. [US/US]; 2225 Lakeridge Road Drive, Grapevine, TX 76051 (US). CADD, Gary, G. [US/US]; 501 Turner Road, Apartment 1111, Grapevine, TX 76051 (US).

- (74) Agent: PRATT, John, S.; Kilpatrick Stockton LLP, Suite 2800, 1100 Peachtree Street, Atlanta, GA 30309 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: VACCINE PRODUCTION USING TRANSPOSON BASED VECTORS

(57) Abstract: The present invention provides novel methods of making proteins, multimeric proteins and antibodies using transposon-based vectors in transgenic individuals that are desirable in producing vaccines. The present invention further provides a method for providing protection against disease by vaccinating an individual with the proteins, multimeric proteins and/or antibodies made in the transgenic individuals of the present invention.

			,
			,
			,
•			

5

15

# 10 VACCINE PRODUCTION USING TRANSPOSON BASED VECTORS

The U.S. Government has certain rights in this invention. The development of this invention was partially funded by the United States Government under a HATCH grant from the United States Department of Agriculture, partially funded by the United States Government with Formula 1433 funds from the United States Department of Agriculture and partially funded by the United States Government under contract DAAD 19-02016 awarded by the Army.

# 20 FIELD OF THE INVENTION

The present invention relates generally to production of proteins, polypeptides and peptides in a transgenic individual, wherein genes encoding the proteins, polypeptides and peptides are operably-linked to signal sequences, or portions of signal sequences. The proteins, polypeptides and peptides are useful as vaccines.

25

30

# BACKGROUND OF THE INVENTION

Diseases caused by numerous biological organisms result in widespread morbidity and mortality throughout the world. Infectious disease is rampant in many locations, and unfortunately, numerous individuals suffer from more than one disease. Immunization of animals and humans with compositions comprising vaccines capable of generating an immune response is one of the only means of preventing or ameliorating the severity of disease.

Numerous diseases threaten not only humans but also animals. Influenza epidemics kill millions each year. Suppy of influenza vaccine is often inadequate for

the needs of numerous individuals. Further, the influenza vaccine is often designed on the best guesses of infectious disease experts concerning the expected antigens and strains for any given year. Different governments and the Centers for Disease Control face a major challenge in redesigning a vaccine and rapidly producing it to supply those in need.

Recent bioterrorism, for example the anthrax attacks, demonstrates the need for a stockpile of antidote. Bioterrorism against animals may also threaten the food supplies of nations. Engineered organisms provide a particular challenge for nations to rapidly adjust their vaccine production to attack newly discovered epitopes in these organisms.

Recent events have demonstrated the vulnerability of the U.S. to attack with biowarfare agents. For many of the agents listed as select agents by the Centers for Disease Control, vaccines are not readily available for humans or domestic livestock. Despite intense research with many of these organisms, the potential for a vaccine in the immediate future is not likely. In addition, treatment with antibiotics may or may not be effective, especially if a bacterium has been engineered to be multi-drug resistant.

Accordingly what is needed in a method for rapid and economical production of proteins, polypeptides, peptides and mixtures thereof that may be used as vaccines to prevent or ameliorate disease.

## SUMMARY OF THE INVENTION

5

10

15

20

25

30

The present invention provides a new, effective and efficient method of producing proteins, polypeptides and peptides (hereinafter generally "proteins") in a transgenic individual that may be used as vaccines. The present invention provides the capability to transform the genetic machinery of individuals in order to produce vaccines. Such transformation may be germ-line transformation and/or non germ-line transformation. The present invention facilitates production of one or more protein, polypeptide and peptide for use in vaccines.

This invention provides polynucleotide cassettes containing at least one gene of interest and one or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. Each of the genes of interest encodes a protein that may be used in the vaccine. One discovery of the present

5

10

15

20

25

30

invention is the use of pro portions of prepro signal sequences to facilitate appropriate processing, expression, and/or formation of proteins in an individual. Preferred pro polynucleotides are provided in SEQ ID NO:1 and SEQ ID NO:2.

Another discovery of the present invention is that cecropin prepro sequences facilitate appropriate processing, expression, and/or formation of desired proteins in an individual. Accordingly, the present invention includes polynucleotide cassettes containing at least one gene of interest operably-linked to a cecropin prepro sequence. Preferred cecropin prepro polynucleotides are provided in SEQ ID NO:3 and SEQ ID NO:4. The present invention also includes polynucleotide cassettes containing at least one gene of interest operably linked to a cecropin prepro polynucleotide, wherein pro sequences are located between the at least one gene of interest when more than one gene of interest is present.

A non-limiting list of the diseases and organisms that may be targets of the vaccines of the present invention follows: Anthrax (Bacillus anthracis), Bacillus anthracis (anthrax), Botulism (Clostridium botulinum toxin), Brucella species (brucellosis), Brucellosis (Brucella species), Burkholderia mallei (glanders), Burkholderia pseudomallei (melioidosis), Chlamydia psittaci (psittacosis), Cholera (Vibrio cholerae), Clostridium botulinum toxin (botulism), Clostridium perfringens (Epsilon toxin), Coxiella burnetii (Q fever), E. coli O157:H7 (Escherichia coli), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of Clostridium perfringens, Escherichia coli O157:H7 (E. coli), Food safety threats (e.g., Salmonella species, Escherichia coli O157:H7, Shigella), Francisella tularensis (tularemia), Glanders (Burkholderia mallei), Melioidosis (Burkholderia pseudomallei), Plague (Yersinia pestis), Psittacosis (Chlamydia psittaci), Q fever (Coxiella burnetii), Ricin toxin from Ricinus communis (castor beans), Rickettsia prowazekii (typhus fever), Salmonella species (salmonellosis), Salmonella Typhi (typhoid fever), Salmonellosis (Salmonella species), Shigellosis), Shigellosis (Shigella), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (Francisella tularensis), Typhoid fever (Salmonella Typhi), Typhus fever (Rickettsia prowazekii), Variola major (smallpox), Vibrio cholerae (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., Vibrio

5

10

15

20

25

30

cholerae, Cryptosporidium parvum), and Yersinia pestis (plague). Other diseases and organisms that may be targets of the vaccines of the present invention are described in Harrison's Principles of Internal Medicine, (Fauci et al. eds, 14<sup>th</sup> edition, McGraw Hill, New York, USA (1998)).

These polynucleotide cassettes are administered to an individual for expression of polypeptide sequences and the formation of a desired protein or multimeric protein for use as a vaccine. Preferably, the individual is an animal from which the multimeric protein can be harvested. In one embodiment, preferred animals are egg-laying or milk-producing animals.

In one embodiment, the egg-laying transgenic animal is an avian. The method of the present invention may be used in avians including Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. Preferably, the egg-laying transgenic animal is a poultry bird. More preferably, the bird is a chicken, turkey, duck, goose or quail. Another preferred bird is a ratite, such as, an emu, an ostrich, a rhea, or a cassowary. Other preferred birds are partridge, pheasant, kiwi, parrot, parakeet, macaw, falcon, eagle, hawk, pigeon, cockatoo, song birds, jay bird, blackbird, finch, warbler, canary, toucan, mynah, or sparrow.

In some embodiments, the polynucleotide cassettes are located within transposon-based vectors that allow for incorporation of the cassettes into the DNA of the individual. The transposon-based vectors of the present invention include a transposase, operably-linked to a first promoter, and a coding sequence for a protein or peptide of interest operably-linked to a second promoter, wherein the coding sequence for the protein or peptide of interest and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the nucleotide at the third base position of each codon is changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of

5

10

15

20

25

30

the transposase gene. In some embodiments, the effective polyA sequence is an avian optimized polyA sequence.

In one embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence comprising ACCATG (SEQ ID NO:5). One example of such a transposon-based vector is the pTnMCS vector (SEQ ID NO:56). In another embodiment the transposon-based vector comprises a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. One example of such a transposon-based vector is the pTnMod vector (SEQ ID NO:57).

Accordingly, it is an object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are antibodies.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are multimeric proteins.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are monoclonal antibodies.

Yet another object of the present invention is to vaccinate individuals, particularly animals and humans, with the vaccines of the present invention in order to prevent or ameliorate disease.

Yet another object of the present invention is to vaccinate individuals, particularly animals and humans, with a vaccine comprising a combination of the proteins or multimeric proteins of the present invention, together with antibodies of the present invention, in order to prevent or ameliorate disease.

Another object of the present invention is to provide proteins, multimeric proteins and antibodies that are useful in the preparation of a medicament that may be used as a vaccine.

An advantage of the present invention is that the proteins are produced more efficiently, in greater quantities and less expensively than taught by prior art methods.

These and other objects, features and advantages of the present invention will become apparent after a review of the following detailed description of the disclosed embodiments and claims.

## 10 BRIEF DESCRIPTION OF THE FIGURES

5

15

20

25

30

Figure 1 depicts schematically a polynucleotide cassette containing two genes of interest operably-linked to two pro polynucleotides, wherein the first pro polynucleotide is a part of a prepro polynucleotide. "Prom" indicates promoter.

Figure 2 depicts schematically a polynucleotide cassette containing polynucleotides encoding for a heavy chain and a light chain of an antibody. "Oval prom" indicates an ovalbumin promoter. The polynucleotide cassette contains pro and prepro sequences and is flanked by insertion sequences (IS) recognized by a transposase.

Figure 3 depicts schematically a polynucleotide cassette containing a cecropin prepro sequence operably-linked to two genes of interest. Between the genes of interest resides a cleavage site indicates by "CS."

Figure 4 depicts schematically a polynucleotide cassette containing two genes of interest, a promoter (prom), a signal sequence (SS) and a cleavage site (CS). The polynucleotide cassette is flanked by insertion sequences (IS) recognized by a transposase.

Figure 5 is a picture of a gel showing partially purified egg white derived from a transgenic avian run under reducing and non-reducing conditions.

# DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a new, effective and efficient method of producing proteins in an individual that may be used as vaccines. Such proteins include polypeptides, peptides and mixtures thereof and are generally called "proteins" throughout the present application. One or more proteins may be made

with the present invention. Antibodies may be made with the present invention and used as vaccines, alone or in conjunction with other proteins and multimeric proteins. Multimeric proteins may be produced with the persent invention and include at least two proteins, wherein the at least two proteins may be the same of different. A multimeric protein includes associated multimeric proteins (two or more associated polypeptides) and multivalent multimeric proteins (a single polypeptide encoded by more than one gene of interest). Expression and/or formation of the multimeric protein in the individual is achieved by administering a polynucleotide cassette containing the gene or genes of interest to the individual. The polynucleotide cassette may additionally contain one or more pro sequences, prepro sequences, cecropin prepro sequences, and/or cleavage site sequences.

5

10

15

20

25

30

One or more genes of interest for incorporation into the polynucleotide cassette are chosen based on the genes that generally code for epitopes of organisms that are believed to be antigenic. Such coding sequences are generally known to one of ordinary skill in the art and can be obtained through GenBank and the published scientific and patent literature. It is within the scope of the present invention to search the genome of organisms to identify coding regios for particular proteins, polypeptides, peptides or epitopes that could serve as epitopes for vaccines. A protein may be made for use in a vaccine against any desired disease or disease producing organism using the method of the present invention. A non-limiting list of the diseases and organisms that may be targets of the vaccines of the present invention follows: Anthrax (Bacillus anthracis), Bacillus anthracis (anthrax), Botulism (Clostridium botulinum toxin), Brucella species (brucellosis), Brucellosis (Brucella species), Burkholderia mallei (glanders), Burkholderia pseudomallei (melioidosis), Chlamydia psittaci (psittacosis), Cholera (Vibrio cholerae), Clostridium botulinum toxin (botulism), Clostridium perfringens (Epsilon toxin), Coxiella burnetii (Q fever), E. coli O157:H7 (Escherichia coli), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of Clostridium perfringens, Escherichia coli O157:H7 (E. coli), Food safety threats (e.g., Salmonella species, Escherichia coli O157:H7, Shigella), Francisella tularensis (tularemia), Glanders (Burkholderia mallei), Melioidosis (Burkholderia pseudomallei), Plague (Yersinia pestis), Psittacosis (Chlamydia psittaci), Q fever (Coxiella burnetii), Ricin toxin from Ricinus communis (castor beans), Rickettsia prowazekii (typhus fever), Salmonella

(salmonellosis), Salmonella Typhi (typhoid fever), Salmonellosis (Salmonella species), Shigella (shigellosis), Shigellosis (Shigella), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (Francisella tularensis), Typhoid fever (Salmonella Typhi), Typhus fever (Rickettsia prowazekii), Variola major (smallpox), Vibrio cholerae (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., Vibrio cholerae, Cryptosporidium parvum), and Yersinia pestis (plague). Other diseases and organisms that may be targets of the vaccines of the present invention are described in Harrison's Principles of Internal Medicine, (Fauci et al. eds, 14<sup>th</sup> edition, McGraw Hill, New York, USA (1998)).

5

10

15

20

25

30

This invention provides polynucleotide cassettes containing at least one gene of interest and one or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. When one gene of interest is present in the polynucleotide cassette, a single protein is made, as opposed to a The present invention provides multimeric protein as described below. polynucleotide cassettes containing two or more genes of interest and two or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro These polynucleotide cassettes are useful for making nucleotide sequence. multimeric proteins. Each of the genes of interest encodes a polypeptide that forms a part of the multimeric protein. These polynucleotide cassettes are administered to an individual for expression of the polypeptide sequences and expression and/or formation of the multimeric protein. Preferably, the individual is an animal from which the proteins or multimeric protein can be harvested. Preferred animals are egglaying or milk-producing animals. In some embodiments, the polynucleotide cassettes are located within transposon-based vectors that allow for incorporation of the cassettes into the DNA of the individual.

The pro polynucleotide sequences operably-linked to the genes of interest include pro portions of prepro polynucleotide sequences commonly associated with polynucleotides encoding proteins secreted from a cell in nature. It may be that the pre polynucleotide sequence functions to direct the resultant protein into the endoplasmic reticulum and the pro sequence is cleaved within the endoplasmic reticulum or Golgi complex of a cell containing the protein. While prepro

5

10

15

20

25

30

polynucleotide sequences are associated with secreted polypeptides in nature, one discovery of the present invention is the use of pro portions of the prepro signal sequences to facilitate appropriate processing, expression, and/or formation of proteins and multimeric proteins, and more particularly, associated multimeric proteins. In the present invention, each gene of interest is operably-linked with a pro polynucleotide sequence. Figure 1 shows schematically one polynucleotide cassette containing two genes of interest, wherein each gene of interest is operably-linked to a pro polynucleotide sequence. The first gene of interest is operably-linked to a pro polynucleotide sequence that is part of a prepro polynucleotide sequence, while the second gene of interest is operably-linked to a pro polynucleotide sequence that is not part of a prepro polynucleotide sequence, but may have been derived from a prepro polynucleotide sequence. Accordingly, the term "pro sequence" encompasses a pro sequence that is part of a prepro sequence and a pro sequence that is not part of a prepro sequence, but may have been derived from a prepro sequence. In preferred embodiments, the most 5' pro polynucleotide sequence in the polynucleotide cassette is a part of a prepro polynucleotide sequence.

Several examples of prepro polynucleotides from which a pro polynucleotide can be derived or be a part of are a cecropin prepro, lysozyme prepro, ovomucin prepro, ovotransferrin prepro, a signal peptide for tumor necrosis factor receptor (SEQ ID NO:6), a signal peptide encoded by a polynucleotide sequence provided in one of SEQ ID NOs:7-54 and a signal peptide provide in SEQ ID NO:55. The prepro or pro polynucleotide can be a cecropin prepro or pro polynucleotide selected from the group consisting of cecropin A1, cecropin A2, cecropin B, cecropin C, cecropin D, cecropin E and cecropin F. In a preferred embodiment, the pro polynucleotide is a cecropin B pro polynucleotide having a sequence shown in SEQ ID NO:1 or SEQ ID NO:2. A preferred prepro polynucleotide is a cecropin B polynucleotide having a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.

Figure 1 provides one embodiment of the invention wherein the polynucleotide cassette includes two genes of interest and two pro polynucleotide sequences arranged in the following order: a prepro polynucleotide, a first gene of interest, a pro polynucleotide, and a second gene of interest. Preferably, the sequences are arranged in the aforementioned order beginning at a 5' end of the polynucleotide cassette. Figure 2 provides a more specific embodiment of the present

invention wherein the first and second genes of interest are polynucleotides encoding antibody heavy and light chains. However, the invention includes polynucleotide cassettes containing at least two genes of interest. Each of the genes of interest is operably-linked to a pro polynucleotide. Each of these pro polynucleotides can be the same, or each can be different. In one embodiment, all of the pro polynucleotides in the polynucleotide cassette are the same and are cecropin pro polynucleotides. The most 5' cecropin pro polynucleotide is preferably a part of a cecropin prepro polynucleotide sequence as shown in Figure 3.

5

10

15

20

25

30

The polynucleotide cassettes of the present invention may be administered to an individual for production of a protein or a multimeric protein in that individual. The discussion that follows is intented to encompass methods for making one protein or a multimeric protein. Accordingly, the present invention includes a method of producing a protein in an individual comprising administering to the individual a polynucleotide cassette comprising one gene of interest encoding the protein, wherein the gene of interest is operably-linked to a pro polynucleotide sequence. The present invention includes a method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide cassette comprising at least two genes of interest, each encoding a part of the multimeric protein, wherein each gene of interest is operably-linked to a pro polynucleotide sequence. The present invention also includes a method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide cassette comprising a cecropin prepro sequence operably-linked to two or more genes of interest, each gene of interest encoding a part of the multimeric protein. This second method does not require the linking of pro polynucleotides to each gene of interest since the use of a cecropin prepro sequence itself in a polynucleotide cassette facilitates processing, expression, and/or formation of multimeric proteins. Polynucleotide cassettes containing the cecropin prepro polynucleotide can contain one or at least two genes of interest. Preferably, the cecropin prepro polynucleotide is located 5' of the one or more genes of interest in the polynucleotide cassette. One exemplary polynucleotide cassette is shown in Figure 3. In a preferred embodiment, the prepro sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4. As shown in Figure 3, the polynucleotide cassettes containing a cecropin prepro polynucleotide preferably contain a cleavage site between each of two genes of interest. Such cleavage site(s)

5

10

15

20

25

30

may be nucleotides encoding any cleavage sites including, but not limited to, an enzymatic cleavage site, a pro polynucleotide, and a photolabile cleavage site, a chemical cleavage site, and a self-splicing cleavage site (i.e., intein). Cleavage sites are discussed in more detail below.

The polynucleotide cassettes of the present invention are particularly suited for production of proteins and/or multimeric proteins in an individual. include both humans and animals. Preferred animals are egg-laying animals and milk-producing animals. As used herein, the term "egg-laying animal" includes all amniotes such as birds, turtles, lizards and monotremes. Monotremes are egg-laying mammals and include the platypus and echidna. The term "bird" or "fowl," as used herein, is defined as a member of the Aves class of animals which are characterized as warm-blooded, egg-laying vertebrates primarily adapted for flying. Avians include, without limitation, Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Coraciformes, Passeriformes. Ralliformes, Cuculiformes. Columbiformes, Galliformes, Anseriformes, and Herodiones. The term "Ratite," as used herein, is defined as a group of flightless, mostly large, running birds comprising several orders and including the emus, ostriches, kiwis, and cassowaries. The term "Psittaciformes", as used herein, includes parrots and refers to a monofamilial order of birds that exhibit zygodactylism and have a strong hooked bill. A "parrot" is defined as any member of the avian family Psittacidae (the single family of the Psittaciformes), distinguished by the short, stout, strongly hooked beak. Preferred avians are poultry birds including chickens, quail, turkeys, geese and ducks. The term "chicken" as used herein denotes chickens used for table egg production, such as egg-type chickens, chickens reared for public meat consumption, or broilers, and chickens reared for both egg and meat production ("dual-purpose" chickens). The term "chicken" also denotes chickens produced by primary breeder companies, or chickens that are the parents, grandparents, great-grandparents, etc. of those chickens reared for public table egg, meat, or table egg and meat consumption.

When the polynucleotide cassettes of the present invention are administered to an egg-laying or milk-producing animal, a transgenic animal containing a polynucleotide cassette is created and the animal produces a transgenic protein or multimeric protein. It is preferred that the resultant protein or multimeric protein is deposited in the egg or in the milk. Various different signal sequences and promoters

may be used to achieve deposition of the protein or multimeric protein in the egg or in the milk and these are described in more detail below. In order to achieve a transgenic animal containing a polynucleotide cassette of the present invention, the polynucleotide cassettes can be administered to the individual with, or contained in, any vector, as naked DNA, or in any delivery construct or solution commonly known to one of ordinary skill in the art. A preferred vector for incorporation of the polynucleotide cassettes into an individual is a transposon-based vector described below.

#### **Definitions**

5

10

15

20

25

30

It is to be understood that as used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for example, reference to "a cell" can mean that at least one cell can be utilized.

The term protein includes polypeptide and peptide and these terms are generally encompassed under the term "protein" and are also used interchangeablt herein. A multimeric protein includes more than one protein, polypeptide or peptide and is defined further below.

The term "antibody" is used interchangeably with the term "immunoglobulin" and is defined herein as a protein synthesized by an animal or a cell of the immune system in response to the presence of a foreign substance commonly referred to as an "antigen" or an "immunoger.". The term antibody includes fragments of antibodies. Antibodies are characterized by specific affinity to a site on the antigen, wherein the site is referred to an "antigenic determinant" or an "epitope". Antigens can be naturally occurring or artificially engineered. Preferred antigens are those from organisms against which a vaccine is desired. Antigens and suspected or known epitopes may be obtained from any organism, including but not limited to those recited herein.

Artificially engineered antigens include but are not limited to small molecules, such as small peptides, attached to haptens such as macromolecules, for example proteins, nucleic acids, or polysaccharides. Artificially designed or engineered variants of naturally occurring antibodies and artificially designed or engineered antibodies not occurring in nature are all included in the current definition. Such variants include conservatively substituted amino acids and other forms of substitution as described in the section concerning proteins and polypeptides.

The term "egg" is defined herein as including a large female sex cell enclosed in a porous, calcarous or leathery shell, produced by birds and reptiles. The term "ovum" is defined as a female gamete, and is also known as an egg. Therefore, egg production in all animals other than birds and reptiles, as used herein, is defined as the production and discharge of an ovum from an ovary, or "ovulation". Accordingly, it is to be understood that the term "egg" as used herein is defined as a large female sex cell enclosed in a porous, calcarous or leathery shell, when a bird or reptile produces it, or it is an ovum when it is produced by all other animals.

5

10

15

20

25

30

The term "gene" is defined herein to include a polynucleotide that includes a coding region for a protein, peptide or polypeptide, with or without intervening sequences such as introns.

The term "multimeric protein" is defined herein to include one or more polypeptides that are associated, or joined, by any means including disulfde bonds. An example of this type of multimeric protein is an antibody that contains both heavy and light chains that are associated by disulfide bonds. These multimeric proteins are referred to herein as "associated multimeric proteins." The term "multimeric protein" also includes a polypeptide that is encoded by more than one gene of interest, such multimeric protein containg desired amino acid sequences encoded by the more than one gene of interest. An example of this type of multimeric protein is a single polypeptide containing a heavy chain polypeptide (first polypeptide of interest) and a light chain polypeptide (second polypeptide of interest). In these embodiments, the different polypeptides of interest may be separated by other polypeptide sequences such as spacer polypeptides and cleavage site polypeptides. These types of multimeric proteins are referred to herein as "multivalent multimeric proteins."

The term "milk-producing animal" refers herein to mammals including, but not limited to, bovine, ovine, porcine, equine, and primate animals. Milk-producing animals include but are not limited to cows, llamas, camels, goats, reindeer, zebu, water buffalo, yak, horses, pigs, rabbits, non-human primates, and humans.

The term "transgenic animal" refers to an animal having at least a portion of the transposon-based vector DNA is incorporated into its DNA. While a transgenic animal includes an animal wherein the transposon-based vector DNA is incorporated into the germline DNA, a transgenic animal also includes an animal having DNA in one or more somatic cells that contain a portion of the transposon-based vector DNA

for any period of time. In a preferred embodiment, a portion of the transposon-based vector comprises a gene of interest. More preferably, the gene of interest is incorporated into the animal's DNA for a period of at least five days, more preferably the laying life of the animal, and most preferably the life of the animal. In a further preferred embodiment, the animal is an avian.

The term "vector" is used interchangeably with the terms "construct", "DNA construct" and "genetic construct" to denote synthetic nucleotide sequences used for manipulation of genetic material, including but not limited to cloning, subcloning, sequencing, or introduction of exogenous genetic material into cells, tissues or organisms, such as birds. It is understood by one skilled in the art that vectors may contain synthetic DNA sequences, naturally occurring DNA sequences, or both. The vectors of the present invention are transposon-based vectors as described herein.

When referring to two nucleotide sequences, one being a regulatory sequence, the term "operably-linked" is defined herein to mean that the two sequences are associated in a manner that allows the regulatory sequence to affect expression of the other nucleotide sequence. It is not required that the operably-linked sequences be directly adjacent to one another with no intervening sequence(s).

The term "regulatory sequence" is defined herein as including promoters, enhancers and other expression control elements such as polyadenylation sequences, matrix attachment sites, insulator regions for expression of multiple genes on a single construct, ribosome entry/attachment sites, introns that are able to enhance expression, and silencers.

## Transposon-Based Vectors

5

10

15

20

25

30

While not wanting to be bound by the following statement, it is believed that the nature of the DNA construct is an important factor in successfully producing transgenic animals. The "standard" types of plasmid and viral vectors that have previously been almost universally used for transgenic work in all species, especially avians, have low efficiencies and may constitute a major reason for the low rates of transformation previously observed. The DNA (or RNA) constructs previously used often do not integrate into the host DNA, or integrate only at low frequencies. Other factors may have also played a part, such as poor entry of the vector into target cells. The present invention provides transposon-based vectors that can be administered to an animal that overcome the prior art problems relating to low transgene integration

5

10

15

20

25

30

frequencies. Two preferred transposon-based vectors of the present invention in which a transposase, gene of interest and other polynucleotide sequences may be introduced are termed pTnMCS (SEQ ID NO:56) and pTnMod (SEQ ID NO:57).

The transposon-based vectors of the present invention produce integration frequencies an order of magnitude greater than has been achieved with previous vectors. More specifically, intratesticular injections performed with a prior art transposon-based vector (described in U.S. Patent No. 5,719,055) resulted in 41% sperm positive roosters whereas intratesticular injections performed with the novel transposon-based vectors of the present invention resulted in 77% sperm positive roosters. Actual frequencies of integration were estimated by either or both comparative strength of the PCR signal from the sperm and histological evaluation of the testes and sperm by quantitative PCR.

The transposon-based vectors of the present invention include a transposase gene operably-linked to a first promoter, and a coding sequence for a desired protein or peptide operably-linked to a second promoter, wherein the coding sequence for the desired protein or peptide and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes one or more of the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of one or more of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. In one embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence comprising ACCATG (SEQ ID NO:5). One example of such a transposon-based vector is the pTnMCS vector (SEQ ID NO:56). In another embodiment the transposon-based vector comprises a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without

changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. One example of such a transposon-based vector is the pTnMod vector (SEQ ID NO:57). The transposon-based vector may additionally or alternatively include one or more of the following Kozak sequences at the 3' end of any promoter, including the promoter operably-linked to the transposase: ACCATGG (SEQ ID NO:58), AAGATGT (SEQ ID NO:59), ACGATGA (SEQ ID NO:60), AAGATGG (SEQ ID NO:61), GACATGA (SEQ ID NO:62), ACCATGA (SEQ ID NO:63), and ACCATGA (SEQ ID NO:64), ACCATGT (SEQ ID NO:65).

#### Transposases and Insertion Sequences

5

10

15

20

25

30

In a further embodiment of the present invention, the transposase found in the transposase-based vector is an altered target site (ATS) transposase and the insertion sequences are those recognized by the ATS transposase. However, the transposase located in the transposase-based vectors is not limited to a modified ATS transposase and can be derived from any transposase. Transposases known in the prior art include those found in AC7, Tn5SEQ1, Tn916, Tn951, Tn1721, Tn 2410, Tn1681, Tn1, Tn2, Tn3, Tn4, Tn5, Tn6, Tn9, Tn10, Tn30, Tn101, Tn903, Tn501, Tn1000 (γδ), Tn1681, Tn2901, AC transposons, Mp transposons, Spm transposons, En transposons, Dotted transposons, Mu transposons, Ds transposons, dSpm transposons and I transposons. According to the present invention, these transposases and their regulatory sequences are modified for improved functioning as follows: a) the addition one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the promoter operably-linked to the transposase; b) a change of one or more of the codons for the first several amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) the addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) the addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene.

Although not wanting to be bound by the following statement, it is believed that the modifications of the first several N-terminal codons of the transposase gene facilitate transcription of the transposase gene, in part, by increasing strand dissociation during transcription. It is preferable that one or more of between

approximately the first 1 to 20, more preferably 3 to 15, and most preferably between 4 to 12 N-terminal codons of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the encoded amino acid. In one embodiment, the first ten N-terminal codons of the transposase gene are modified in this manner. It is also preferred that the transposase contain mutations that make it less specific for preferred insertion sites and thus increases the rate of transgene insertion as discussed in U.S. Patent No. 5,719,055.

In some embodiments, the transposon-based vectors are optimized for expression in a particular host by changing the methylation patterns of the vector DNA. For example, prokaryotic methylation may be reduced by using a methylation deficient organism for production of the transposon-based vector. The transposon-based vectors may also be methylated to resemble eukaryotic DNA for expression in a eukaryotic host.

10

15

20

25

30

Transposases and insertion sequences from other analogous eukaryotic transposon-based vectors that can also be modified and used are, for example, the Drosophila P element derived vectors disclosed in U.S. Patent No. 6,291,243; the Drosophila mariner element described in Sherman et al. (1998); or the sleeping beauty transposon. See also Hackett et al. (1999); D. Lampe et al., 1999. Proc. Natl. Acad. Sci. USA, 96:11428-11433; S. Fischer et al., 2001. Proc. Natl. Acad. Sci. USA, 98:6759-6764; L. Zagoraiou et al., 2001. Proc. Natl. Acad. Sci. USA, 98:11474-11478; and D. Berg et al. (Eds.), Mobile DNA, Amer. Soc. Microbiol. (Washington, D.C., 1989). However, it should be noted that bacterial transposon-based elements are preferred, as there is less likelihood that a eukaryotic transposase in the recipient species will recognize prokaryotic insertion sequences bracketing the transgene.

Many transposases recognize different insertion sequences, and therefore, it is to be understood that a transposase-based vector will contain insertion sequences recognized by the particular transposase also found in the transposase-based vector. In a preferred embodiment of the invention, the insertion sequences have been shortened to about 70 base pairs in length as compared to those found in wild-type transposons that typically contain insertion sequences of well over 100 base pairs.

While the examples provided below incorporate a "cut and insert" Tn10 based vector that is destroyed following the insertion event, the present invention also encompasses the use of a "rolling replication" type transposon-based vector. Use of a

PCT/US2003/041269 WO 2004/067743

rolling replication type transposon allows multiple copies of the transposon/transgene to be made from a single transgene construct and the copies inserted. This type of transposon-based system thereby provides for insertion of multiple copies of a transgene into a single genome. A rolling replication type transposon-based vector may be preferred when the promoter operably-linked to gene of interest is endogenous to the host cell and present in a high copy number or highly expressed. However, use of a rolling replication system may require tight control to limit the insertion events to non-lethal levels. Tn1, Tn2, Tn3, Tn4, Tn5, Tn9, Tn21, Tn501, Tn551, Tn951, Tn1721, Tn2410 and Tn2603 are examples of a rolling replication type transposon, although Tn5 could be both a rolling replication and a cut and insert type transposon.

# Stop Codons and PolyA Sequences

5

10

15

20

25

30

In one embodiment, the transposon-based vector contains two stop codons operably-linked to the transposase and/or to the gene of interest. In an alternate embodiment, one stop codon of UAA or UGA is operably linked to the transposase and/or to the gene of interest.

As used herein an "effective polyA sequence" refers to either a synthetic or non-synthetic sequence that contains multiple and sequential nucleotides containing an adenine base (an A polynucleotide string) and that increases expression of the gene to which it is operably-linked. A polyA sequence may be operably-linked to any gene in the transposon-based vector including, but not limited to, a transposase gene and a gene of interest. A preferred polyA sequence is optimized for use in the host animal or human. In one embodiment, the polyA sequence is optimized for use in an avian species and more specifically, a chicken. An avian optimized polyA sequence generally contains a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string and thereby separate the stop codon from the A In one embodiment of the present invention, the polyA polynucleotide string. sequence comprises a conalbumin polyA sequence as provided in SEQ ID NO:66 and as taken from GenBank accession # Y00407, base pairs 10651-11058. In another embodiment, the polyA sequence comprises a synthetic polynucleotide sequence shown in SEQ ID NO:67. In yet another embodiment, the polyA sequence comprises an avian optimized polyA sequence provided in SEQ ID NO:68. A chicken optimized

polyA sequence may also have a reduced amount of CT repeats as compared to a synthetic polyA sequence.

It is a suprising discovery of the present invention that such an avian optimized poly A sequence increases expression of a polynucleotide to which it is operably-linked in an avian as compared to a non-avian optimized polyA sequence. Accordingly, the present invention includes methods of or increasing incorporation of a gene of interest wherein the gene of interest resides in a transposon-based vector containing a transposase gene and wherein the transposase gene is operably linked to an avian optimized polyA sequence. The present invention also includes methods of increasing expression of a gene of interest in an avian that includes administering a gene of interest to the avian, wherein the gene of interest is operably-linked to an avian optimized polyA sequence. An avian optimized polyA nucleotide string is defined herein as a polynucleotide containing an A polynucleotide string and a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A The present invention further provides transposon-based polynucleotide string. vectors containing a gene of interest or transposase gene operably linked to an avian optimized polyA sequence.

#### Promoters and Enhancers

10

15

20

25

30

The first promoter cperably-linked to the transposase gene and the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. Constitutive promoters include, but are not limited to, immediate early cytomegalovirus (CMV) promoter, herpes simplex virus 1 (HSV1) immediate early promoter, SV40 promoter, lysozyme promoter, early and late CMV promoters, early and late HSV promoters,  $\beta$ -actin promoter, tubulin promoter, Rous-Sarcoma virus (RSV) promoter, and heat-shock protein (HSP) promoter. Inducible promoters include tissue-specific promoters, developmentally-regulated promoters and chemically inducible promoters. Examples of tissue-specific promoters include the glucose 6 phosphate (G6P) promoter, vitellogenin promoter, ovalbumin promoter, ovomucoid promoter, conalbumin promoter, ovotransferrin promoter, prolactin promoter, kidney uromodulin promoter, and placental lactogen promoter. In one embodiment, the vitellogenin promoter includes a polynucleotide sequence of SEQ ID NO:69. The G6P promoter sequence may be deduced from a rat G6P gene

untranslated upstream region provided in GenBank accession number U57552.1. Examples of developmentally-regulated promoters include the homeobox promoters and several hormone induced promoters. Examples of chemically inducible promoters include reproductive hormone induced promoters and antibiotic inducible promoters such as the tetracycline inducible promoter and the zinc-inducible metallothionine promoter.

5

10

15

20

25

30

Other inducible promoter systems include the Lac operator repressor system inducible by IPTG (isopropyl beta-D-thiogalactoside) (Cronin, A. et al. 2001. Genes and Development, v. 15), ecdysone-based inducible systems (Hoppe, U. C. et al. 2000. Mol. Ther. 1:159-164); estrogen-based inducible systems (Braselmann, S. et al. 1993. Proc. Natl. Acad. Sci. 90:1657-1661); progesterone-based inducible systems using a chimeric regulator, GLVP, which is a hybrid protein consisting of the GALA binding domain and the herpes simplex virus transcriptional activation domain, VP16, and a truncated form of the human progesterone receptor that retains the ability to bind ligand and can be turned on by RU486 (Wang, et al. 1994. Proc. Natl. Acad. Sci. 91:8180-8184); CID-based inducible systems using chemical inducers of dimerization (CIDs) to regulate gene expression, such as a system wherein rapamycin induces dimerization of the cellular proteins FKBP12 and FRAP (Belshaw, P. J. et al. 1996. J. Chem. Biol. 3:731-738; Fan, L. et al. 1999. Hum. Gene Ther. 10:2273-2285; Shariat, S.F. et al. 2001. Cancer Res. 61:2562-2571; Spencer, D.M. 1996. Curr. Biol. 6:839-847). Chemical substances that activate the chemically inducible promoters can be administered to the animal containing the transgene of interest via any method known to those of skill in the art.

Other examples of cell or tissue-specific and constitutive promoters include but are not limited to smooth-muscle SM22 promoter, including chimeric SM22alpha/telokin promoters (Hoggatt A.M. et al., 2002. Circ Res. 91(12):1151-9); ubiquitin C promoter (Biochim Biophys Acta, 2003. Jan. 3;1625(1):52-63); Hsf2 promoter; murine COMP (cartilage oligomeric matrix protein) promoter; early B cell-specific mb-1 promoter (Sigvardsson M., et al., 2002. Mol. Cell Biol. 22(24):8539-51); prostate specific antigen (PSA) promoter (Yoshimura I. et al., 2002, J. Urol. 168(6):2659-64); exorh promoter and pineal expression-promoting element (Asaoka Y., et al., 2002. Proc. Natl. Acad. Sci. 99(24):15456-61); neural and liver ceramidase gene promoters (Okino N. et al., 2002. Biochem. Biophys. Res. Commun.

299(1):160-6); PSP94 gene promoter/enhancer (Gabril M.Y. et al., 2002. Gene Ther. 9(23):1589-99); promoter of the human FAT/CD36 gene (Kuriki C., et al., 2002. Biol. Pharm. Bull. 25(11):1476-8); VL30 promoter (Staplin W.R. et al., 2002. Blood October 24, 2002); and IL-10 promoter (Brenner S., et al., 2002. J. Biol. Chem. December 18, 2002).

5

10

15

20

25

30

Examples of avian promoters include, but are not limited to, promoters controlling expression of egg white proteins, such as ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, ovostatin (ovomacroglobin), cystatin, avidin, thiamine-binding protein, glutamyl aminopeptidase minor glycoprotein 1, minor glycoprotein 2; and promoters controlling expression of egg-yolk proteins, such as vitellogenin, very lowdensity lipoproteins, low density lipoprotein, cobalamin-binding protein, riboflavinbinding protein, biotin-binding protein (Awade, 1996. Z. Lebensm. Unters. Forsch. 202:1-14). An advantage of using the vitellogenin promoter is that it is active during the egg-laying stage of an animal's life-cycle, which allows for the production of the protein of interest to be temporally connected to the import of the protein of interest into the egg yolk when the protein of interest is equipped with an appropriate targeting sequence. In some embodiments, the avian promoter is an oviduct-specific promoter. As used herein, the term "oviduct-specific promoter" includes, but is not limited to, ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin (ovomacroglobin) promoters.

Liver-specific promoters of the present invention include, but are not limited to, the following promoters, vitellogenin promoter, G6P promoter, cholesterol-7-alpha-hydroxylase (CYP7A) promoter, phenylalanine hydroxylase (PAH) promoter, protein C gene promoter, insulin-like growth factor I (IGF-I) promoter, bilirubin UDP-glucuronosyltransferase promoter, aldolase B promoter, furin promoter, metallothioneine promoter, albumin promoter, and insulin promoter.

Also included in the present invention are promoters that can be used to target expression of a protein of interest into the milk of a milk-producing animal including, but not limited to,  $\beta$  lactoglobin promoter, whey acidic protein promoter, lactalbumin promoter and casein promoter.

Promoters associated with cells of the immune system may also be used. Acute phase promoters such as interleukin (IL)-1 and IL-2 may be employed. Promoters for heavy and light chain Ig may also be employed. The promoters of the T cell receptor components CD4 and CD8, B cell promoters and the promoters of CR2 (complement receptor type 2) may also be employed. Immune system promoters are preferably used when the desired protein is an antibody protein.

5

10

15

20

25

30

Also included in this invention are modified promoters/enhancers wherein elements of a single promoter are duplicated, modified, or otherwise changed. In one embodiment, steroid hormone-binding domains of the ovalbumin promoter are moved from about -6.5 kb to within approximately the first 1000 base pairs of the gene of interest. Modifying an existing promoter with promoter/enhancer elements not found naturally in the promoter, as well as building an entirely synthetic promoter, or drawing promoter/enhancer elements from various genes together on a non-natural backbone, are all encompassed by the current invention.

Accordingly, it is to be understood that the promoters contained within the transposon-based vectors of the present invention may be entire promoter sequences or fragments of promoter sequences. For example, in one embodiment, the promoter operably linked to a gene of interest is an approximately 900 base pair fragment of a chicken ovalbumin promoter (SEQ ID NO:70). The constitutive and inducible promoters contained within the transposon-based vectors may also be modified by the addition of one or more modified Kozak sequences of ACCATG (SEQ ID NO:5).

As indicated above, the present invention includes transposon-based vectors containing one or more enhancers. These enhancers may or may not be operably-linked to their native promoter and may be located at any distance from their operably-linked promoter. A promoter operably-linked to an enhancer is referred to herein as an "enhanced promoter." The enhancers contained within the transposon-based vectors are preferably enhancers found in birds, and more preferably, an ovalbumin enhancer, but are not limited to these types of enhancers. In one embodiment, an approximately 675 base pair enhancer element of an ovalbumin promoter is cloned upstream of an ovalbumin promoter with 300 base pairs of spacer DNA separating the enhancer and promoter. In one embodiment, the enhancer used as a part of the present invention comprises base pairs 1-675 of a Chicken Ovalbumin

enhancer from GenBank accession #S82527.1. The polynucleotide sequence of this enhancer is provided in SEQ ID NO:71.

Also included in some of the transposon-based vectors of the present invention are cap sites and fragments of cap sites. In one embodiment, approximately 50 base pairs of a 5' untranslated region wherein the capsite resides are added on the 3' end of an enhanced promoter or promoter. An exemplary 5' untranslated region is provided in SEQ ID NO:72. A putative cap-site residing in this 5' untranslated region preferably comprises the polynucleotide sequence provided in SEQ ID NO:73.

5

10

15

20

25

30

In one embodiment of the present invention, the first promoter operably-linked to the transposase gene is a constitutive promoter and the second promoter operably-linked to the gene of interest is a tissue-specific promoter. In the second embodiment, use of the first constitutive promoter allows for constitutive activation of the transposase gene and incorporation of the gene of interest into virtually all cell types, including the germline of the recipient animal. Although the gene of interest is incorporated into the germline generally, the gene of interest is only expressed in a tissue-specific manner. A transposon-based vector having a constitutive promoter operably-linked to the transposase gene can be administered by any route, and in one embodiment, the vector is administered to an ovary or to an artery leading to the ovary. In another embodiment, the vector is administered into the lumen of the oviduct or into an artery supplying the oviduct.

It should be noted that cell- or tissue-specific expression as described herein does not require a complete absence of expression in cells or tissues other than the preferred cell or tissue. Instead, "cell-specific" or "tissue-specific" expression refers to a majority of the expression of a particular gene of interest in the preferred cell or tissue, respectively.

When incorporation of the gene of interest into the germline is not preferred, the first promoter operably-linked to the transposase gene can be a tissue-specific promoter. For example, transfection of a transposon-based vector containing a transposase gene operably-linked to an oviduct specific promoter such as the ovalbumin promoter provides for activation of the transposase gene and incorporation of the gene of interest in the cells of the oviduct but not into the germline and other cells generally. In this embodiment, the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. In a preferred

embodiment, both the first promoter and the second promoter are an ovalbumin promoter. In embodiments wherein tissue-specific expression or incorporation is desired, it is preferred that the transposon-based vector is administered directly to the tissue of interest or to an artery leading to the tissue of interest. In a preferred embodiment, the tissue of interest is the oviduct and administration is achieved by direct injection into the lumen of the oviduct or an artery leading to the oviduct. In a further preferred embodiment, administration is achieved by direct injection into the lumen of the magnum or the infundibulum of the oviduct.

Accordingly, cell specific promoters may be used to enhance transcription in selected tissues. In birds, for example, promoters that are found in cells of the fallopian tube, such as ovalbumin, conalbumin, ovomucoid and/or lysozyme, are used in the vectors to ensure transcription of the gene of interest in the epithelial cells and tubular gland cells of the fallopian tube, leading to synthesis of the desired protein encoded by the gene and deposition into the egg white. In mammals, promoters specific for the epithelial cells of the alveoli of the mammary gland, such as prolactin, insulin, beta lactoglobin, whey acidic protein, lactalbumin, casein, and/or placental lactogen, are used in the design of vectors used for transfection of these cells for the production of desired proteins for deposition into the milk. In liver cells, the G6P promoter may be employed to drive transcription of the gene of interest for protein production. Proteins made in the liver of birds may be delivered to the egg yolk.

In order to achieve higher or more efficient expression of the transposase gene, the promoter and other regulatory sequences operably-linked to the transposase gene may be those derived from the host. These host specific regulatory sequences can be tissue specific as described above or can be of a constitutive nature. For example, an avian actin promoter and its associated polyA sequence can be operably-linked to a transposase in a transposase-based vector for transfection into an avian. Examples of other host specific promoters that could be operably-linked to the transposase include the myosin and DNA or RNA polymerase promoters.

## Directing Sequences

5

10

15

20

25

30

In some embodiments of the present invention, the gene of interest is operably-linked to a directing sequence or a sequence that provides proper conformation to the desired protein encoded by the gene of interest. As used herein, the term "directing sequence" refers to both signal sequences and targeting sequences.

An egg directing sequence includes, but is not limited to, an ovomucoid signal sequence, an ovalbumin signal sequence, a cecropin pre pro signal sequence, and a vitellogenin targeting sequence. The term "signal sequence" refers to an amino acid sequence, or the polynucleotide sequence that encodes the amino acid sequence, a portion or the entirety of which directs the protein to which it is linked to the endoplasmic reticulum in a eukaryote, and more preferably the translocational pores in the endoplasmic reticulum, or the plasma membrane in a prokaryote, or mitochondria, such as for the purpose of gene therapy for mitochondrial diseases. Signal and targeting sequences can be used to direct a desired protein into, for example, the milk, when the transposon-based vectors are administered to a milk-producing animal.

Signal sequences can also be used to direct a desired protein into, for example, a secretory pathway for incorporation into the egg yolk or the egg white, when the transposon-based vectors are administered to a bird or other egg-laying animal. The present invention also includes a gene of interest operably-linked to a second gene containing a signal sequence. An example of such an embodiment is wherein the gene of interest is operably-linked to the ovalbumin gene that contains an ovalbumin signal sequence. Other signal sequences that can be included in the transposon-based vectors include, but are not limited to the ovotransferrin and lysozyme signal sequences. In one embodiment, the signal sequence is an ovalbumin signal sequence including a sequence shown in SEQ ID NO:74. In another embodiment, the signal sequence is a shortened ovalbumin signal sequence including a sequence shown in SEQ ID NO:75 or SEQ ID NO:76.

As also used herein, the term "targeting sequence" refers to an amino acid sequence, or the polynucleotide sequence encoding the amino acid sequence, which amino acid sequence is recognized by a receptor located on the exterior of a cell. Binding of the receptor to the targeting sequence results in uptake of the protein or peptide operably-linked to the targeting sequence by the cell. One example of a targeting sequence is a vitellogenin targeting sequence that is recognized by a vitellogenin receptor (or the low density lipoprotein receptor) on the exterior of an oocyte. In one embodiment, the vitellogenin targeting sequence includes the polynucleotide sequence of SEQ ID NO:77. In another embodiment, the vitellogenin targeting sequence includes all or part of the vitellogenin gene. Other targeting

sequences include VLDL and Apo E, which are also capable of binding the vitellogenin receptor. Since the ApoE protein is not endogenously expressed in birds, its presence may be used advantageously to identify birds carrying the transposon-based vectors of the present invention.

## 5 Genes of Interest

10

15

20

25

30

The genes of interest in the polynucleotide cassette can be any gene, and preferably are genes that encode proteins or portions of multimeric proteins that may be used in vaccines. Such genes may be coding regions for desired epitopes in infectious organisms, including but not limited to viral, bacterial, protozoal, fungal, and parasitic organisms. Such genes are known to one of ordinary skill in the art and may be obtained from GenBank and from scientific and patent literature. A gene of interest may contain modifications of the codons for the first several N-terminal amino acids of the gene of interest, wherein the third base of each codon is changed to an A or a T without changing the corresponding amino acid.

In one embodiment, the genes of interest are antibody genes or portions of antibody genes. Figure 2 shows a schematic drawing of a polynucleotide cassette containing an antibody heavy chain and an antibody light chain as two genes of interest. Antibodies used in or encoded by the polynucleotide cassettes of the present invention include, but are not limited to, IgG, IgM, IgA, IgD, IgE, IgY, lambda chains, kappa chains, bi-specific antibodies, and fragments thereof; scFv fragments, Fc fragments, and Fab fragments as well as dimeric, trimeric and oligomeric forms of antibody fragments. Desired antibodies include, but are not limited to, naturally occurring antibodies, human antibodies, humanized antibodies, autoantibodies and hybrid antibodies. Preferred antibodies include those useful as vaccines against selected diseases. Genes encoding modified versions of naturally occurring antibodies or fragments thereof and genes encoding artificially designed antibodies or fragments thereof may be incorporated into the transposon-based vectors of the present invention. Desired antibodies also include antibodies with the ability to bind specific ligands, for example, antibodies against proteins associated with diseases, particularly infectious diseases. Accordingly, the present invention encompasses a polynucleotide cassette as described herein containing one or more genes encoding a heavy immunoglobulin (Ig) chain and a light Ig chain.

5

10

15

20

25

30

Antibodies that may be produced using the present invention include, but are not limited to, antibodies for providing passive immunity to an animal or a human against an infectious disease or a toxic agent. The antibodies prepared using the methods of the present invention may also be designed to possess specific labels that may be detected through means known to one of ordinary skill in the art. For example, antibodies may be labeled with a fluorescent label attached that may be detected following exposure to specific wavelengths. Such labeled antibodies may be primary antibodies directed to a specific antigen, for example, rhodamine-labeled rabbit anti-growth hormone, or may be labeled secondary antibodies, such as fluorescein-labeled goat-anti chicken IgG. Such labeled antibodies are known to one of ordinary skill in the art. The antibodies may also be designed to possess specific sequences useful for purification through means known to one of ordinary skill in the art. Labels useful for attachment to antibodies are also known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Proteins and multimeric proteins made according to the present invention may also be labeled with these methods. Antibodies, proteins and multimeric proteins produced with the present invention may be used as laboratory reagents for numerous applications including radioimmunoassay, western blots, dot blots, ELISA, immunoaffinity columns and other procedures requiring antibodies as known to one of ordinary skill in the art. Such antibodies include primary antibodies, secondary antibodies and tertiary antibodies, which may be labeled or unlabeled.

Additional antibodies that may be made with the practice of the present invention include, but are not limited to, primary antibodies, secondary antibodies, designer antibodies, anti-protein antibodies, anti-peptide antibodies, anti-DNA antibodies, anti-RNA antibodies, antibodies against non-natural antigens, anti-venom antibodies, antibodies directed against epitopes associated with infectious disease, including, anti-viral, anti-bacterial, anti-protozoal, anti-fungal, anti-parasitic, anti-receptor, anti-lipid, anti-phospholipid, anti-growth factor, anti-cytokine, anti-monokine, anti-idiotype, and anti-accessory (presentation) protein antibodies. Antibodies made with the present invention, as well as light chains or heavy chains, may also be used to inhibit enzyme activity.

5

10

15

Antibodies that may be produced using the present invention include, but are not limited to, antibodies made against the following proteins: Bovine  $\gamma$ -Globulin, Serum; Bovine IgG, Plasma; Chicken  $\gamma$ -Globulin, Serum; Human  $\gamma$ -Globulin, Serum; Human IgA, Plasma; Human IgA1, Myeloma; Human IgA2, Myeloma; Human IgA2, Plasma; Human IgD, Plasma; Human IgE, Myeloma; Human IgG, Plasma; Human IgG, Fab Fragment, Plasma; Human IgG, F(ab')2 Fragment, Plasma; Human IgG, Fc Fragment, Plasma; Human IgG1, Myeloma; Human IgG2, Myeloma; Human IgG3, Myeloma; Human IgG4, Myeloma; Human IgM, Myeloma; Human IgM, Plasma; Human Immunoglobulin, Light Chains  $\kappa$  and  $\lambda$  Plasma; Mouse  $\gamma$ -Globulin, Serum; Mouse IgG, Serum; Mouse IgM, Myeloma; Rabbit  $\gamma$ -Globulin, Serum; Rabbit IgG, Plasma; and Rat  $\gamma$ -Globulin, Serum. In one embodiment, the transposon-based vector comprises the coding sequence of light and heavy chains of a murine monoclonal antibody that shows specificity for human seminoprotein (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

A further non-limiting list of antibodies that recognize other antibodies and that may be produced using the present invention is as follows: Anti-Chicken IgG, heavy (H) & light (L) Chain Specific (Sheep); Anti-Goat γ-Globulin (Donkey); Anti-Goat IgG, Fc Fragment Specific (Rabbit); Anti-Guinea Pig  $\gamma$ -Globulin (Goat); Anti-Human Ig, Light Chain, Type  $\kappa$  Specific; Anti-Human Ig, Light Chain, Type  $\lambda$ 20 Specific; Anti-Human IgA, α-Chain Specific (Goat); Anti-Human IgA, Fab Fragment Specific; Anti-Human IgA, Fc Fragment Specific; Anti-Human IgA, Secretory; Anti-Human IgE, e-Chain Specific (Goat); Anti-Human IgE, Fc Fragment Specific; Anti-Human IgG, Fc Fragment Specific (Goat); Anti-Human IgG, γ-Chain Specific (Goat); Anti-Human IgG, Fc Fragment Specific; Anti-Human IgG, Fd Fragment Specific; 25 Anti-Human IgG, H & L Chain Specific (Goat); Anti-Human IgG1, Fc Fragment Specific; Anti-Human IgG2, Fc Fragment Specific; Anti-Human IgG2, Fd Fragment Specific; Anti-Human IgG3, Hinge Specific; Anti-Human IgG4, Fc Fragment Specific; Anti-Human IgM, Fc Fragment Specific; Anti-Human IgM, µ-Chain Specific; Anti-Mouse IgE, ε-Chain Specific; Anti-Mouse γ-Globulin (Goat); Anti-Mouse IgG, γ-30 Chain Specific (Goat); Anti-Mouse IgG, \( \gamma\)-Chain Specific (Goat) F(ab')2 Fragment; Anti-Mouse IgG, H & L Chain Specific (Goat); Anti-Mouse IgM, µ-Chain Specific (Goat); Anti-Mouse IgM, H & L Chain Specific (Goat); Anti-Rabbit γ-Globulin

(Goat); Anti-Rabbit IgG, Fc Fragment Specific (Goat); Anti-Rabbit IgG, H & L Chain Specific (Goat); Anti-Rat γ-Globulin (Goat); Anti-Rat IgG, H & L Chain Specific; Anti-Rhesus Monkey γ-Globulin (Goat); and, Anti-Sheep IgG, H & L Chain Specific.

Antibodies may be made against peptides such as p146 (SEQ ID NO:78 amino acid sequence, SEQ ID NO:79, nucleotide sequence), and other lytic peptides or peptides associated with disease, particularly infectious disease.

5

10

15

20

25

30

The following is yet another non-limiting of antibodies that can be produced by the methods of present invention: abciximab (ReoPro), abciximab anti-platelet aggregation monoclonal antibody, anti-CD11a (hu1124), anti-CD18 antibody, anti-CD20 antibody, anti-cytomegalovirus (CMV) antibody, anti-digoxin antibody, antihepatitis B antibody, anti-HER-2 antibody, anti-idiotype antibody to GD3 glycolipid, anti-IgE antibody, anti-IL-2R antibody, antimetastatic cancer antibody (mAb 17-1A), anti-rabies antibody, anti-respiratory syncytial virus (RSV) antibody, anti-Rh antibody, anti-TCR, anti-TNF antibody, anti-VEGF antibody and Fab fragment thereof, rattlesnake venom antibody, black widow spider venom antibody, coral snake venom antibody, antibody against very late antigen-4 (VLA-4), C225 humanized antibody to EGF receptor, chimeric (human & mouse) antibody against TNFa, antibody directed against GPIIb/IIIa receptor on human platelets, gamma globulin, anti-hepatitis B immunoglobulin, human anti-D immunoglobulin, human antibodies against S aureus, human tetanus immunoglobulin, humanized antibody against the epidermal growth receptor-2, humanized antibody against the  $\alpha$  subunit of the interleukin-2 receptor, humanized antibody CTLA4IG, humanized antibody to the IL-2 R α-chain, humanized anti-CD40-ligand monoclonal antibody (5c8), humanized mAb against the epidermal growth receptor-2, humanized mAb to rous sarcoma virus, humanized recombinant antibody (IgG1k) against respiratory syncytial virus (RSV), lymphocyte immunoglobulin (anti-thymocyte antibody), lymphocyte immunoglobulin, mAb against factor VII, MDX-210 bi-specific antibody against HER-2, MDX-22, MDX-220 bi-specific antibody against TAG-72 on tumors, MDX-33 antibody to FcγR1 receptor, MDX-447 bi-specific antibody against EGF receptor, MDX-447 bispecific humanized antibody to EGF receptor, MDX-RA immunotoxin (ricin A linked) antibody, Medi-507 antibody (humanized form of BTI-322) against CD2 receptor on T-cells, monoclonal antibody LDP-02, muromonab-CD3(OKT3)

5

10

15

20

25

30

antibody, OKT3 ("muromomab-CD3") antibody, PRO 542 antibody, ReoPro ("abciximab") antibody, and TNF-IgG fusion protein.

Antibodies may also be made, in accordance with the present invention, against the following non-limiting list of organisms associated with disease and against peptides and/or proteins or other epitopes from these organisms: Anthrax (Bacillus anthracis), Bacillus anthracis (anthrax), Botulism (Clostridium botulinum toxin). Brucella species (brucellosis), Brucellosis (Brucella species), Burkholderia mallei (glanders), Burkholderia pseudomallei (melioidosis), Chlamydia psittaci (psittacosis), Cholera (Vibrio cholerae), Clostridium botulinum toxin (botulism), Clostridium perfringens (Epsilon toxin), Coxiella burnetii (Q fever), E. coli O157:H7 (Escherichia coli), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of Clostridium perfringens, Escherichia coli O157:H7 (E. coli), Food safety threats (e.g., Salmonella species, Escherichia coli O157:H7, Shigella), Francisella tularensis (tularemia), Glanders (Burkholderia mallei), Melioidosis (Burkholderia pseudomallei), Plague (Yersinia pestis), Psittacosis (Chlamydia psittaci), Q fever (Coxiella burnetii), Ricin toxin from Ricinus communis (castor beans), Rickettsia prowazekii (typhus fever), Salmonella species (salmonellosis), Salmonella Typhi (typhoid fever), Salmonellosis (Salmonella species), Shigella (shigellosis), Shigellosis (Shigella), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (Francisella tularensis), Typhoid fever (Salmonella Typhi), Typhus feyer (Rickettsia prowazekii), Variola major (smallpox), Vibrio cholerae (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., Vibrio cholerae, Cryptosporidium parvum), and Yersinia pestis (plague).

Another non-limiting list of the antibodies that may be produced using the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs San Carlos CA, SIGMA, St.Louis, MO www.sigma-aldrich.com, Cappel ICN, Irvine, California, www.icnbiomed.com, and Calbiochem, La Jolla, California, www.calbiochem.com, which are all incorporated herein by reference in their entirety. The polynucleotide sequences encoding these antibodies

5

10

15

20

25

30

may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the antibody polynucleotide sequence by choosing the codons that encode for each amino acid in the desired antibody.

As an example of protein production using the method of the present invention which may be applied to the production of proteins useful in vaccines, the gene of interest is a proinsulin gene and the desired molecule is insulin. Proinsulin consists of three parts: a C-peptide and two strands of amino acids (the alpha and beta chains) that later become linked together to form the insulin molecule. In these embodiments, proinsulin is expressed in the oviduct tubular gland cells and then deposited in the egg white. One example of a proinsulin polynucleotide sequence is shown in SEQ ID NO:80, wherein the C-peptide cleavage site spans from Arg at position 31 to Arg at position 65.

Further included in the present invention are genes of interet that encode proteins and peptides synthesized by the immune system including those synthesized by the thymus, lymph nodes, spleen, and the gastrointestinal associated lymph tissues (GALT) system. The immune system proteins and peptides proteins that can be made in transgenic animals using the polynucleotide cassettes of the present invention include, but are not limited to, alpha-interferon, beta-interferon, gamma-interferon, alpha-interferon A, alpha-interferon 1, G-CSF, GM-CSF, interlukin-1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, TNF-α, and TNF-β. Other cytokines included in the present invention include cardiotrophin, stromal cell derived factor, macrophage derived chemokine (MDC), melanoma growth stimulatory activity (MGSA), macrophage inflammatory proteins 1 alpha (MIP-1 alpha), 2, 3 alpha, 3 beta, 4 and 5.

Genes encoding lytic peptides such as p146 are also included in the genes of interest of the present invention. In one embodiment, the p146 peptide comprises an amino acid sequence of SEQ ID NO:78. The present invention also encompasses a polynucleotide cassette comprising a p146 nucleic acid having a sequence of SEQ ID NO:79.

Enzymes are another class of proteins that may be encoded by the polynucleotide cassettes of the present invention. Such enzymes include but are not limited to adenosine deaminase, alpha-galactosidase, cellulase, collagenase, dnaseI,

hyaluronidase, lactase, L-asparaginase, pancreatin, papain, streptokinase B, subtilisin, superoxide dismutase, thrombin, trypsin, urokinase, fibrinolysin, glucocerebrosidase and plasminogen activator. In some embodiments wherein the enzyme could have deleterious effects, additional amino acids and a protease cleavage site are added to the carboxy end of the enzyme of interest in order to prevent expression of a functional enzyme. Subsequent digestion of the enzyme with a protease results in activation of the enzyme. Preferred are enzymes associate with disease processes and the function of infectious organisms.

A non-limiting list of the peptides and proteins that may be encoded by the polynucleotide cassettes of the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA, (St.Louis, MO www.sigma-aldrich.com), Cappel ICN (Irvine, California, www.icnbiomed.com), and Calbiochem (La Jolla, California, www.calbiochem.com). The polynucleotide sequences encoding these proteins and peptides of interest may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons that encode for each amino acid in the desired protein or peptide.

Other desired proteins that may be encoded by the polynucleotide cassettes of the present invention include bacitracin, polymixin b, vancomycin, cyclosporine, anti-RSV antibody, alpha-1 antitrypsin (AAT), anti-cytomegalovirus antibody, antihepatitis antibody, anti-inhibitor coagulant complex, anti-rabies antibody, anti-Rh(D) antibody, adenosine deaminase, anti-digoxin antibody, antivenin crotalidae (rattlesnake venom antibody), antivenin latrodectus (black widow spider venom antibody), antivenin micrurus (coral snake venom antibody), aprotinin, corticotropin (ACTH), diphtheria antitoxin, lymphocyte immune globulin (anti-thymocyte antibody), protamine, thyrotropin, capreomycin,  $\alpha$ -galactosidase, gramicidin, streptokinase, tetanus toxoid, tyrothricin, IGF-1, proteins of varicella vaccine, anti-TNF antibody, anti-IL-2r antibody, anti-HER-2 antibody, OKT3 ("muromonab-CD3") antibody, TNF-IgG fusion protein, ReoPro ("abciximab") antibody, ACTH fragment 1-24, desmopressin, gonadotropin-releasing hormone, histrelin, leuprolide, lypressin, nafarelin, peptide that binds GPIIb/GPIIIa on platelets (integrilin),

5

10

15

20

25

30

goserelin, capreomycin, colistin, anti-respiratory syncytial virus, lymphocyte immune globulin (Thymoglovin, Atgam), panorex, alpha-antitrypsin, botulinin, lung surfactant protein, tumor necrosis receptor-IgG fusion protein (enbrel), gonadorelin, proteins of influenza vaccine, proteins of rotavirus vaccine, proteins of haemophilus b conjugate vaccine, proteins of poliovirus vaccine, proteins of pneumococcal conjugate vaccine, proteins of meningococcal C vaccine, proteins of influenza vaccine, megakaryocyte growth and development factor (MGDF), neuroimmunophilin ligand-A (NIL-A), brain-derived neurotrophic factor (BDNF), glial cell line-derived neurotrophic factor (GDNF), leptin (native), leptin B, leptin C, IL-1RA (interleukin-1RA), R-568, novel erythropoiesis-stimulating protein (NESP), humanized mAb to rous sarcoma virus (MEDI-493), glutamyl-tryptophan dipeptide IM862, LFA-3TIP immunosuppressive, humanized anti-CD40-ligand monoclonal antibody (5c8), gelsonin enzyme, tissue factor pathway inhibitor (TFPI), proteins of meningitis B vaccine, antimetastatic cancer antibody (mAb 17-1A), chimeric (human & mouse) mAb against TNFa, mAb against factor VII, relaxin, capreomycin, glycopeptide (LY333328), recombinant human activated protein C (rhAPC), humanized mAb against the epidermal growth receptor-2, altepase, anti-CD20 antigen, C2B8 antibody, insulin-like growth factor-1, atrial natriuretic peptide (anaritide), tenectaplase, anti-CD11a antibody (hu 1124), anti-CD18 antibody, mAb LDP-02, anti-VEGF antibody, fab fragment of anti-VEGF Ab, APO2 ligand (tumor necrosis factor-related apoptosis-inducing ligand), rTGF-β (transforming growth factor- $\beta$ ), alpha-antitrypsin, ananain (a pineapple enzyme), humanized mAb CTLA4IG, PRO 542 (mAb), D2E7 (mAb), calf intestine alkaline phosphatase, α-L-iduronidase, α-L-galactosidase (humanglutamic acid decarboxylase, acid sphingomyelinase, bone morphogenetic protein-2 (rhBMP-2), proteins of HIV vaccine, T cell receptor (TCR) peptide vaccine, TCR peptides, V beta 3 and V beta 13.1. (IR502), (IR501), BI 1050/1272 mAb against very late antigen-4 (VLA-4), C225 humanized mAb to EGF receptor, anti-idiotype antibody to GD3 glycolipid, antibacterial peptide against H. pylori, MDX-447 bispecific humanized mAb to EGF receptor, anti-cytomegalovirus (CMV), Medi-491 B19 parvovirus vaccine, humanized recombinant mAb (IgG1k) against respiratory syncytial virus (RSV), urinary tract infection vaccine (against "pili" on Escherechia coli strains), proteins of lyme disease vaccine against B. burgdorferi protein (DbpA), proteins of Medi-501 human papilloma virus-11 vaccine (HPV), Streptococcus pneumoniae vaccine, Medi-507

mAb (humanized form of BTI-322) against CD2 receptor on T-cells, MDX-33 mAb to Fc $\gamma$ R1 receptor, MDX-RA immunotoxin (ricin A linked) mAb, MDX-210 bispecific mAb against HER-2, MDX-447 bi-specific mAb against EGF receptor, MDX-22, MDX-220 bi-specific mAb against TAG-72 on tumors, colony-stimulating factor (CSF) (molgramostim), humanized mAb to the IL-2 R  $\alpha$ -chain (basiliximab), mAb to IgE (IGE 025A), myelin basic protein-altered peptide (MSP771A), humanized mAb against the epidermal growth receptor-2, humanized mAb against the  $\alpha$  subunit of the interleukin-2 receptor, low molecular weight heparin, anti-hemophillic factor, and bactericidal/permeability-increasing protein (r-BPI).

The following paragraph describes specific examples of epitopes and antibodies associated with specific diseases that may be made as the proteins, multimeric proteins of the present invention. Such epitopes may also be used in antibody production. Anthrax: PA (protective antigen) mediates entry of toxin's lethal factor (LF) into cell, therefore PA is binding domain. There are at least three different binding sites on PA where LF can bind. Site I epitope IG3PA63, site II 2D3PA, 2D5PA, 10D2PA, and site III 3B6PA, 14B7PA, 10E10PA63. Botulism: BoNT (botulism neurotoxin), binding domains within neurotoxin A (BoNT/A) derived from segments of heavy chain and encoded proteins: H455-661, H1150-1289. Accession number X73844. BoNT/A in infants: Accession number X73423.

Tularemia: mAbs against LPS of francisella tularensis: FT14, FT2F11. OMP specific mAbs react to IgGl subclass of *F. tularensis*. Yersenia pestis: causes pneumonic and bubonic plague. LCrV: accession number M26405 (*Y. pestis*), LCrV: accession number M57893 (*Y. pseudotuberculosis*). Caf-1 encodes capsular antigen fraction 1 (F1) accession number X61996. Shigella: epitopes 2C32E6, 4D64B9 (reactive to *S. flexneri* and *S. boydii*). 5E45D8 – *S. flexneri*. 4B33D10, 1B52F10 – all species of Shigella. Shigella monoclonal antibodies 2C32E6, 4D64B9, 5E45D8 (reactive with S. Felxneri). 4B33D10 and 1B52F1 (species of Shigella) and 1B52F10 all strains (see Rahman & Stimson, Hybridoma, 20:85-90, 2001). See Venkatesan et al., PNAS 85:9317-21 (1988) accession number J04117 S. flexnari). IpaD between amino acid residues 14 and 44 of Shigella flexnari. IpaD is against the unique region of this protein exposed on the surface of the pathogen (Turbyfill et al., Infect. Immunity 66:1999-2006 (1998). See Muller-Loennies et al., J. Biol. Chem. 278: 25618-27 (2003) for epitopes related to Shigella. LcrV, Yersinia pestis (accession.

5

10

15

20

25

30

number M26405), Yersinia pseudotuberculosis (accessionnumber M57893, Yersinia enterocolitica. Fusion of IcrV and the structural gene for protein A. One protective epitope resides internally between amino acids 168 and 275 (Motin et al., Infect. Immunity 62:41920-201 (1994). For additional epitopes on Yersinia pastis see accession number M26405 LcrV (Price et al., J. Bacteriol. 171: 5646-5653 (1989). For additional epitopes on Yersinia pastis see cafl encoding cafl capsular antigen fraction 1 (F1) accession number X61996 in Galyov et al., FEBS Lett. 277: 230-2 (1990). 12G12 SLP-M common specific epitope of the Brucella S-LPS. S-LPS of B. melitensis Rev1, S-LPS smooth lipopolysaccharide, (Weynants et al., Clin. Diagn. Lab. Immunol. 3:309-14 (1996). Smooth Brucella species - two lipopolysacchiaride (LPSs (A and M) produce monoclonal antibodies to the specific A- or M-LPS epitopes. 12AE6, and Mab 33.1.5, s-LPSA=12AE6, LPSM=33.1.5, BmE10-5, 12G12 O side chain = BRU-38 (Douglas et al., J. Clin. Microbiol. 26: 1353-6 (1988). BmE10-5 Mantigen of Brucella melitensis 16M, B. melitensis, B. abortus, B. suis and B. neotomae (Vizcaino et al., Res. Microbiol. 143: 513-8 (1992). BRU-38, O side chain of B. abortus (Roop et al., J. Clin. Microbiol. 25: 2090-3 (1987). The OMPspecific monoclonal antibodies Bhatti et al., Hybridoma 12: 197-202 (1993). Mabs FT14 and FT2F11, Francisella tularensis (Fulop et al., J. Clin. Microbiol. 29: 1407-12 (1991).

Preferred proteins and multimeric proteins to be made with the present invention include those associated with disease processes and those associated with antigenic epitopes of infectious organisms.

The proteins and multimeric proteins made using the present invention may be labeled using labels and techniques known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Some of these labels may be genetically engineered into the polynucleotide sequence for the expression of the selected multimeric protein. The peptides and proteins may also have label-incorporation "handles" incorporated to allow labeling of an otherwise difficult or impossible to label multimeric protein.

It is to be understood that the various classes of desired peptides and proteins, as well as specific peptides and proteins described in this section may be modified as

described below by inserting selected codons for desired amino acid substitutions into the gene incorporated into the transgenic animal.

The present invention further encompasses the use of inhibitory molecules to inhibit endogenous (i.e., non-vector) protein production. These inhibitory molecules include antisense nucleic acids, siRNA and inhibitory proteins. Preferred inhibitory molecules are those associated with causation or inhibition of disease processes. In a preferred embodiment, the endogenous protein whose expression is inhibited is an egg white protein including, but not limited to ovalbumin, ovotransferrin, and ovomucin ovomucoid, ovoinhibitor, cystatin, ovostatin, lysozyme, ovoglobulin G2, ovoglobulin G3, avidin, and thiamin binding protein. In one embodiment, a polynucleotide cassette containing an ovalbumin DNA sequence, that upon transcription forms a double stranded RNA molecule, is transfected into an animal such as a bird and the bird's production of endogenous ovalbumin protein is reduced by the interference RNA mechanism (RNAi). In other embodiments, a polynucleotide cassette encodes an inhibitory RNA molecule that inhibits the expression of more than one egg white protein. Additionally, inducible knockouts or knockdowns of the endogenous protein may be created to achieve a reduction or inhibition of endogenous protein production. Endogenous egg white production can be inhibited in an avian at any time, but is preferably inhibited preceding, or immediately preceding, the harvest of eggs.

## 20 Modified Desired Proteins and Peptides

5

10

15

25

30

The present invention may be used for the production of proteins and multimeric proteins. "Proteins", "peptides," "polypeptides" and "oligopeptides" are chains of amino acids (typically L-amino acids) whose alpha carbons are linked through peptide bonds formed by a condensation reaction between the carboxyl group of the alpha carbon of one amino acid and the amino group of the alpha carbon of another amino acid. The terminal amino acid at one end of the chain (i.e., the amino terminal) has a free amino group, while the terminal amino acid at the other end of the chain (i.e., the carboxy terminal) has a free carboxyl group. As such, the term "amino terminus" (abbreviated N-terminus) refers to the free alpha-amino group on the amino acid at the amino terminal of the protein, or to the alpha-amino group (imino group when participating in a peptide bond) of an amino acid at any other location within the protein. Similarly, the term "carboxy terminus" (abbreviated C-terminus) refers to

the free carboxyl group on the amino acid at the carboxy terminus of a protein, or to the carboxyl group of an amino acid at any other location within the protein.

Typically, the amino acids making up a protein are numbered in order, starting at the amino terminal and increasing in the direction toward the carboxy terminal of the protein. Thus, when one amino acid is said to "follow" another, that amino acid is positioned closer to the carboxy terminal of the protein than the preceding amino acid.

The term "residue" is used herein to refer to an amino acid (D or L) or an amino acid mimetic that is incorporated into a protein by an amide bond. As such, the amino acid may be a naturally occurring amino acid or, unless otherwise limited, may encompass known analogs of natural amino acids that function in a manner similar to the naturally occurring amino acids (i.e., amino acid mimetics). Moreover, an amide bond mimetic includes peptide backbone modifications well known to those skilled in the art.

Furthermore, one of skill will recognize that, as mentioned above, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than about 5%, more typically less than about 1%) in an encoded sequence are conservatively modified variations where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);

5

10

15

20

30

- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

A conservative substitution is a substitution in which the substituting amino acid (naturally occurring or modified) is structurally related to the amino acid being substituted, i.e., has about the same size and electronic properties as the amino acid being substituted. Thus, the substituting amino acid would have the same or a similar functional group in the side chain as the original amino acid. A "conservative substitution" also refers to utilizing a substituting amino acid which is identical to the

5

10

15

20

25

amino acid being substituted except that a functional group in the side chain is protected with a suitable protecting group.

Suitable protecting groups are described in Green and Wuts, "Protecting Groups in Organic Synthesis", John Wiley and Sons, Chapters 5 and 7, 1991, the teachings of which are incorporated herein by reference. Preferred protecting groups are those which facilitate transport of the peptide through membranes, for example, by reducing the hydrophilicity and increasing the lipophilicity of the peptide, and which can be cleaved, either by hydrolysis or enzymatically (Ditter et al., 1968. J. Pharm. Sci. 57:783; Ditter et al., 1968. J. Pharm. Sci. 57:828; Ditter et al., 1969. J. Pharm. Sci. 58:557; King et al., 1987. Biochemistry 26:2294; Lindberg et al., 1989. Drug Metabolism and Disposition 17:311; Tunek et al., 1988. Biochem. Pharm. 37:3867; Anderson et al., 1985 Arch. Biochem. Biophys. 239:538; and Singhal et al., 1987. FASEB J. 1:220). Suitable hydroxyl protecting groups include ester, carbonate and carbamate protecting groups. Suitable amine protecting groups include acyl groups and alkoxy or aryloxy carbonyl groups, as described above for N-terminal protecting groups. Suitable carboxylic acid protecting groups include aliphatic, benzyl and aryl esters, as described below for C-terminal protecting groups. In one embodiment, the carboxylic acid group in the side chain of one or more glutamic acid or aspartic acid residues in a peptide of the present invention is protected, preferably as a methyl, ethyl, benzyl or substituted bonzyl ester, more preferably as a benzyl ester.

Provided below are groups of naturally occurring and modified amino acids in which each amino acid in a group has similar electronic and steric properties. Thus, a conservative substitution can be made by substituting an amino acid with another amino acid from the same group. It is to be understood that these groups are non-limiting, i.e. that there are additional modified amino acids which could be included in each group.

- Group I includes leucine, isoleucine, valine, methionine and modified amino acids having the following side chains: ethyl, n-propyl n-butyl. Preferably, Group I includes leucine, isoleucine, valine and methionine.
- Group II includes glycine, alanine, valine and a modified amino acid having an ethyl side chain. Preferably, Group II includes glycine and alanine.
  - Group III includes phenylalanine, phenylglycine, tyrosine, tryptophan, cyclohexylmethyl glycine, and modified amino residues having substituted

5

10

25

30

benzyl or phenyl side chains. Preferred substituents include one or more of the following: halogen, methyl, ethyl, nitro, —NH<sub>2</sub>, methoxy, ethoxy and — CN. Preferably, Group III includes phenylalanine, tyrosine and tryptophan.

Group IV includes glutamic acid, aspartic acid, a substituted or unsubstituted aliphatic, aromatic or benzylic ester of glutamic or aspartic acid (e.g., methyl, ethyl, n-propyl iso-propyl, cyclohexyl, benzyl or substituted benzyl), glutamine, asparagine, —CO—NH— alkylated glutamine or asparagines (e.g., methyl, ethyl, n-propyl and iso-propyl) and modified amino acids having the side chain —(CH<sub>2</sub>)<sub>3</sub>—COOH, an ester thereof (substituted or unsubstituted aliphatic, aromatic or benzylic ester), an amide thereof and a substituted or unsubstituted N-alkylated amide thereof. Preferably, Group IV includes glutamic acid, aspartic acid, methyl aspartate, ethyl aspartate, benzyl aspartate and methyl glutamate, ethyl glutamate and benzyl glutamate, glutamine and asparagine.

15 Group V includes histidine, lysine, ornithine, arginine, N-nitroarginine, β-cycloarginine, γ-hydroxyarginine, N-amidinocitruline and 2-amino-4-guanidinobutanoic acid, homologs of lysine, homologs of arginine and homologs of ornithine. Preferably, Group V includes histidine, lysine, arginine and ornithine. A homolog of an amino acid includes from 1 to about 3 additional or subtracted methylene units in the side chain.

Group VI includes serine, threonine, cysteine and modified amino acids having C1-C5 straight or branched alkyl side chains substituted with —OH or —SH, for example, —CH<sub>2</sub>CH<sub>2</sub>OH, —CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>OH or -CH<sub>2</sub>CH<sub>2</sub>OHCH<sub>3</sub>. Preferably, Group VI includes serine, cysteine or threonine.

In another aspect, suitable substitutions for amino acid residues include "severe" substitutions. A "severe substitution" is a substitution in which the substituting amino acid (naturally occurring or modified) has significantly different size and/or electronic properties compared with the amino acid being substituted. Thus, the side chain of the substituting amino acid can be significantly larger (or smaller) than the side chain of the amino acid being substituted and/or can have functional groups with significantly different electronic properties than the amino acid being substituted. Examples of severe substitutions of this type include the substitution of phenylalanine or cyclohexylmethyl glycine for alanine, isoleucine for

glycine, a D amino acid for the corresponding L amino acid, or —NH—CH[(—CH<sub>2</sub>)<sub>5</sub>—COOH]—CO— for aspartic acid. Alternatively, a functional group may be added to the side chain, deleted from the side chain or exchanged with another functional group. Examples of severe substitutions of this type include adding of valine, leucine or isoleucine, exchanging the carboxylic acid in the side chain of aspartic acid or glutamic acid with an amine, or deleting the amine group in the side chain of lysine or ornithine. In yet another alternative, the side chain of the substituting amino acid can have significantly different steric and electronic properties that the functional group of the amino acid being substituted. Examples of such modifications include tryptophan for glycine, lysine for aspartic acid and — (CH<sub>2</sub>)<sub>4</sub>COOH for the side chain of serine. These examples are not meant to be limiting.

In another embodiment, for example in the synthesis of a peptide 26 amino acids in length, the individual amino acids may be substituted according in the following manner:

AA<sub>1</sub> is serine, glycine, alanine, cysteine or threonine;

AA2 is alanine, threonine, glycine, cysteine or serine;

AA<sub>3</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;

AA<sub>4</sub> is proline, leucine, valine, isoleucine or methionine;

5

10

15

20

AA<sub>5</sub> is tryptophan, alanine, phenylalanine, tyrosine or glycine;

AA6 is serine, glycine, alanine, cysteine or threonine;

AA<sub>7</sub> is proline, leucine, valine, isoleucine or methionine;

25 AA<sub>8</sub> is alanine, threonine, glycine, cysteine or serine;

AA9 is alanine, threonine, glycine, cysteine or serine;

AA<sub>10</sub> is leucine, isoleucine, methionine or valine;

AA<sub>11</sub> is serine, glycine, alanine, cysteine or threonine;

AA<sub>12</sub> is leucine, isoleucine, methionine or valine;

30 AA<sub>13</sub> is leucine, isoleucine, methionine or valine;

AA<sub>14</sub> is glutamine, glutamic acid, aspartic acid, asparagine, or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

 $AA_{15}$  is arginine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxy-arginine, N-amidinocitruline or 2-amino-4-guanidino-butanoic acid

AA<sub>16</sub> is proline, leucine, valine, isoleucine or methionine;

AA<sub>17</sub> is serine, glycine, alanine, cysteine or threonine;

5 AA<sub>18</sub> is glutamic acid, aspartic acid, asparagine, glutamine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

AA<sub>19</sub> is aspartic acid, asparagine, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

10 AA<sub>20</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, β-cycloarginine, γ-hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;

AA<sub>21</sub> is alanine, threonine, glycine, cysteine or serine;

AA<sub>22</sub> is alanine, threonine, glycine, cysteine or serine;

20

25

30

15 AA<sub>23</sub> is histidine, serine, threonine, cysteine, lysine or ornithine;

AA<sub>24</sub> is threonine, aspartic acid, serine, glutamic acid or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

AA<sub>25</sub> is asparagine, aspartic acid,, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid; and

AA<sub>26</sub> is cysteine, histidine, serine, threonine, lysine or ornithine.

It is to be understood that these amino acid substitutions may be made for longer or shorter peptides than the 26 mer in the preceding example above, and for proteins.

In one embodiment of the present invention, codons for the first several N-terminal amino acids of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the gene of interest are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. In one embodiment, the first ten N-terminal codons of the gene of interest are modified in this manner.

When several desired proteins, protein fragments or peptides are encoded in the gene of interest to be incorporated into the genome, as with the multivalent multimeric proteins, one of skill in the art will appreciate that the proteins, protein fragments or peptides may be separated by a spacer molecule such as, for example, a peptide, consisting of one or more amino acids. Generally, the spacer will have no specific biological activity other than to join the desired proteins, protein fragments or peptides together, or to preserve some minimum distance or other spatial relationship between them. However, the constituent amino acids of the spacer may be selected to influence some property of the molecule such as the folding, net charge, or hydrophobicity. The spacer may also be contained within a nucleotide sequence with a purification handle or be flanked by proteolytic cleavage sites.

Such polypeptide spacers may have from about 1 to about 100 amino acids, preferably 3 to 20 amino acids, and more preferably 4-15 amino acids. The spacers in a polypeptide are independently chosen, but are preferably all the same. The spacers should allow for flexibility of movement in space and are therefore typically rich in small amino acids, for example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are (Gly-Pro-Gly-Gly)<sub>x</sub> (SEQ ID NO:81) and (Gly<sub>4</sub>-Ser)<sub>y</sub>, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include

(Gly-Pro-Gly-Gly)3

5

10

15

20

30

SEQ ID NO:82 Gly Pro Gly Gly Gly Pro Gly Gly Pro Gly Gly (Gly<sub>4</sub>-Ser)<sub>3</sub>

One example of a multivalent multimeric protein containin a spacer is leutinizing hormone (LH), normally made as separate alpha and beta chains, made as a single polypeptide as described in Galet et. al., Mol. Cell Endocrinology, 2001, 174 (1-2):31-40. Production of a multimeric protein may thus be simplified using a spacer sequence that may or may not contain cleavage sites. In the case of an

immunoglobulin, for example, a heavy and light chain may be synthesized as a single polypeptide using a spacer sequence with protease sites native to the transgenic animal so as to make, upon processing, a heavy and light chain combination in close association, facilitating the addition of a similar heavy and light chain to produce the native immunoglobulin. In this model, the removal of the spacer sequence may or may not be required. Other multimeric proteins may be made in bioengineered organisms in a similar fashion.

5

10

15

20

25

30

Nucleotide sequences encoding for the production of residues which may be useful in purification of the expressed recombinant protein may also be built into the vector. Such sequences are known in the art and include the glutathione binding domain from glutathione S-transferase, polylysine, hexa-histidine or other cationic amino acids, thioredoxin, hemagglutinin antigen and maltose binding protein.

Additionally, nucleotide sequences may be inserted into the gene of interest to be incorporated so that the protein or peptide can also include from one to about six amino acids that create signals for proteolytic cleavage. In this manner, if a gene is designed to make one or more peptides or proteins of interest in the transgenic animal, specific nucleotide sequences encoding for amino acids recognized by enzymes may be incorporated into the gene to facilitate cleavage of the large protein or peptide sequence into desired peptides or proteins or both. For example, nucleotides encoding a proteolytic cleavage site can be introduced into the gene of interest so that a signal sequence can be cleaved from a protein or peptide encoded by the gene of interest. Nucleotide sequences encoding other amino acid sequences which display pH sensitivity, chemical sensitivity or photolability may also be added to the vector to facilitate separation of the signal sequence from the peptide or protein of interest.

Proteolytic cleavage sites include cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Chemical cleavage sites are also included in the defintion of cleavage site as used herein. Chemical cleavage sites include, but are not limited to, site cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid. Self-splicing cleavage sites such as inteins are also included in the present invention.

In some embodiments, one or more cleavage sites are incorporated into a polynucleotide cassette containing multiple genes of interest. Figure 4 depicts one example of a polynucleotide cassette containing two genes of interest containing a cleavage site between them. The genes of interest may encode different proteins or peptides, the same protein or peptide, or modified versions of the same protein or peptide. While Figure 4 shows a polynucleotide cassette containing two genes of interest, the present invention encompasses a polynucleotide cassette containing any number of genes of interest. The cleavage site located between the genes of interest can encode any amino acid sequence that is cleaved by any means. As mentioned above, the cleavage site can encode an amino acid sequence cleaved by a protease, a chemical reaction, can be a photolabile site, or can be a pro polynucleotide.

The present invention includes a polynucleotide cassette that encodes a repetitive polypeptide chain in which two or more peptides, polypeptides or proteins, designated as P in the structural formulae presented below, are each separated by a peptide spacer or cleavage site designated as B. A polypeptide multivalent ligand, also called a multivalent protein, is a form of a multimeric protein encoded by the polynucleotide cassettes of the present invention, and is represented by structural formulae (I, II and III). Each peptide or protein is connected to another peptide or protein through a peptide bond, to a linker group, to a spacer, or to a cleavage site. Each peptide, polypeptide or protein may be the same or different and each linker, spacer, cleavage site or covalent bond is independently chosen.

A "polypeptide multivalent protein" is a multiple repeat polypeptide chain in which two or more peptides P are each separated by a peptide linker group, a spacer or a cleavage site. A polypeptide multivalent ligand is represented by structural formulae II and III.

$$I = B - (L - P)n$$

5

10

15

20

25

30

wherein B is a peptide spacer or cleavage site, n is an integer from 2 to about 20, each L is a covalent bond, a linking group or cleavage site which may be present or absent, and each P is a peptide having from about 4 to about 200 amino acid residues.

wherein m is an integer from 0 to about 20.

III 
$$Pa - (B)n - Pa$$

wherein n is an integer from 1 to 20, preferably 2 to 10, more preferably 3 to 7, further wherein a is 1.

Other examples of multivalent proteins include the following:

$$IV$$
  $P_y$ — $L_x$ — $B_n$ — $L_x$ — $P_y$ 

10

15

20

25

30

$$V$$
  $P_y$ — $B$ — $P_y$ 

In the preceding structural formulae IV and V of polypeptide multivalent ligands encoded by a polynucleotide cassette of the present invention, each P is a peptide having from about 4 to about 200 amino acid residues, y is 1, x is an integer from 1 to 3, and n is an integer from 1 to 20, preferably 2 to 10, more preferably 3 to 7. Each B is a peptide spacer or cleavage site comprised of at least 2 amino acids or a cleavage site. Each peptide P and each B are independently chosen and may be the same or different.

Suitable linkers (L) are groups that can connect peptides and proteins to each other. In one example, the linker is an oligopeptide of from about 1 to about 10 amino acids consisting of amino acids with inert side chains. Suitable oligopeptides include polyglycine, polyserine, polyproline, polyalanine and oligopeptides consisting of alanyl and/or serinyl and/or prolinyl and/or glycyl amino acid residues. m in structural formula II is an integer from 0 to about 20.

The peptides, polypeptides and proteins in a multivalent protein can be connected to each other by covalent bonds, linker groups, spacers, cleavage groups or a combination thereof. The linking groups can be the same or different.

A polypeptide spacer shown in structural formula (II) is a peptide having from about 5 to about 40 amino acid residues. The spacers in a polypeptide multivalent ligand are independently chosen, and may be the same or different. The spacers should allow for flexibility of movement in space for the flanking peptides. polypeptides and proteins P, and are therefore typically rich in small amino acids, for

example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are (Gly-Pro-Gly-Gly)<sub>x</sub> (SEQ ID NO:81) and (Gly<sub>4</sub>-Ser)<sub>y</sub>, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include (Gly<sub>4</sub>-Ser)<sub>3</sub> (SEQ ID NO:82). Spacers can also include from one to about four amino acids that create a signal for proteolytic cleavage.

5

10

15

20

25

30

In another embodiment of the present invention, a TAG sequence is linked to a gene of interest. The TAG sequence serves three purposes: 1) it allows free rotation of the peptide or protein to be isolated so there is no interference from the native protein or signal sequence, i.e. vitellogenin, 2) it provides a "purification handle" to isolate the protein using affinity purification, and 3) it includes a cleavage site to remove the desired protein from the signal and purification sequences. Accordingly, as used herein, a TAG sequence includes a spacer sequence, a purification handle and a cleavage site. The spacer sequences in the TAG proteins contain one or more repeats shown in SEQ ID NO:85. A preferred spacer sequence comprises the sequence provided in SEQ ID NO:86. One example of a purification handle is the gp41 hairpin loop from HIV I. Exemplary gp41 polynucleotide and polypeptide sequences are provided in SEQ ID NO:87 and SEQ ID NO:88, respectively. However, it should be understood that any antigenic region, or otherwise associative regions such as avidin/biotin, may be used as a purification handle, including any antigenic region of gp41. Preferred purification handles are those that elicit highly specific antibodies. Additionally, the cleavage site can be any protein cleavage site known to one of ordinary skill in the art and includes an enterokinase cleavage site comprising the Asp Asp Asp Asp Lys sequence (SEQ ID NO:89) and a furin cleavage In one embodiment of the present invention, the TAG sequence comprises a site. polynucleotide sequence of SEQ ID NO:90.

#### Methods of Administering Polynucleotide Cassettes

In addition to the polynucleotide cassettes described above, the present invention also includes methods of administering the polynucleotide cassettes to an animal, methods of producing a transgenic animal wherein a gene of interest is incorporated into the germline of the animal and methods of producing a transgenic

animal wherein a gene of interest is incorporated into cells other than the germline cells of the animal. The polynucleotide cassettes may reside in any vector or delivery solution when administered or may be naked DNA. In one embodiment, a transposon-based vector containing the polynucleotide cassette between two insertion sequences recognized by a transposase is administered to an animal. The polynucleotide cassettes of the present invention may be administered to an animal via any method known to those of skill in the art, including, but not limited to, intraembryonic, intratesticular, intraoviduct, intraovarian, into the duct system of the mammary gland, intraperitoneal, intraarterial, intravenous, topical, oral, nasal, and pronuclear injection methods of administration, or any combination thereof. The polynucleotide cassettes may also be administered within the lumen of an organ, into an organ, into a body cavity, into the cerebrospinal fluid, through the urinary system, through the genitourinary system, through the reproductive system, or through any route to reach the desired cells.

The polynucleotide cassettes may be delivered through the vascular system to be distributed to the cells supplied by that vessel. For example, the compositions may be placed in the artery supplying the ovary or supplying the fallopian tube to transfect cells in those tissues. In this manner, follicles could be transfected to create a germline transgenic animal. Alternatively, supplying the compositions through the artery leading to the oviduct would preferably transfect the tubular gland and epithelial cells. Such transfected cells could manufacture a desired protein or peptide for deposition in the egg white. Administration of the compositions through the portal vein would target uptake and transformation of hepatic cells. Administration through the urethra and into the bladder would target the transitional epithelium of the bladder. Administration through the vagina and cervix would target the lining of the uterus and the epithelial cells of the fallopian tube. Administration through the internal mammary artery or through the duct system of the mammary gland would transfect secretory cells of the lactating mammary gland to perform a desired function, such as to synthesize and secrete a desired protein or peptide into the milk.

The polynucleotide cassettes may be administered in a single administration, multiple administrations, continuously, or intermittently. The polynucleotide cassettes may be administered by injection, via a catheter, an osmotic mini-pump or any other method. In some embodiments, a polynucleotide cassette is administered to

5

10

15

20

25

30

an animal in multiple administrations, each administration containing the polynucleotide cassette and a different transfecting reagent.

In a preferred embodiment, the animal is an egg-laying animal, and more preferably, an avian. In one embodiment, between approximately 1 and 150 µg, 1 and 100 µg, 1 and 50 µg, preferably between 1 and 20 µg, and more preferably between 5 and 10 µgof a transposon-based vector containing the polynucleotide cassette is administered to the oviduct of a bird. In a chicken, it is preferred that between approximately 1 and 100 µg, or 5 and 50 µg are administered. In a quail, it is preferred that between approximately 5 and 10 µg are administered. Optimal ranges depending upon the type of bird and the bird's stage of sexual maturity. Intraoviduct administration of the transposon-based vectors of the present invention result in a PCR positive signal in the oviduct tissue, whereas intravascular administration results in a PCR positive signal in the liver. In other embodiments, the polynucleotide cassettes is administration may also be combined with any methods for facilitating transfection, including without limitation, electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

The transposon-based vectors may be administered to the animal at any point during the lifetime of the animal, however, it is preferable that the vectors are administered prior to the animal reaching sexual maturity. The transposon-based vectors are preferably administered to a chicken oviduct between approximately 14 and 16 weeks of age and to a quail oviduct between approximately 5 and 10 weeks of age, more preferably 5 and 8 weeks of age, and most preferably between 5 and 6 weeks of age, when standard poultry rearing practices are used. The vectors may be administered at earlier ages when exogenous hormones are used to induce early sexual maturation in the bird. In some embodiments, the transposon-based vector is administered to an animal's oviduct following an increase in proliferation of the oviduct epithelial cells and/or the tubular gland cells. Such an increase in proliferation normally follows an influx of reproductive hormones in the area of the oviduct. When the animal is an avian, the transposon-based vector is administered to the avian's oviduct following an increase in proliferation of the oviduct epithelial cells and before the avian begins to produce egg white constituents.

The present invention also includes a method of intraembryonic administration of a transposon-based vector containing a polynucleotide cassette to an avian embryo comprising the following steps: 1) incubating an egg on its side at room temperature for two hours to allow the embryo contained therein to move to top dead center (TDC); 2) drilling a hole through the shell without penetrating the underlying shell membrane; 3) injecting the embryo with the transposon-based vector in solution; 4) sealing the hole in the egg; and 5) placing the egg in an incubator for hatching. Administration of the transposon-based vector can occur anytime between immediately after egg lay (when the embryo is at Stage X) and hatching. Preferably, the transposon-based vector is administered between 1 and 7 days after egg lay, more preferably between 1 and 2 days after egg lay. The transposon-based vectors may be introduced into the embryo in amounts ranging from about 5.0 µg to 10 pg, preferably 1.0 µg to 100 pg. Additionally, the transposon-based vector solution volume may be between approximately 1 µl to 75 µl in quail and between approximately 1 µl to 500 µl in chicken.

The present invention also includes a method of intratesticular administration of a transposon-based vector containing a polynucleotide cassette including injecting a bird with a composition comprising the transposon-based vector, an appropriate carrier and an appropriate transfection reagent. In one embodiment, the bird is injected before sexual maturity, preferably between approximately 4-14 weeks, more preferably between approximately 6-14 weeks and most preferably between 8-12 weeks old. In another embodiment, a mature bird is injected with a transposon-based vector an appropriate carrier and an appropriate transfection reagent. The mature bird may be any type of bird, but in one example the mature bird is a quail.

A bird is preferably injected prior to the development of the blood-testis barrier, which thereby facilitates entry of the transposon-based vector into the seminiferous tubules and transfection of the spermatogonia or other germline cells. At and between the ages of 4, 6, 8, 10, 12, and 14 weeks, it is believed that the testes of chickens are likely to be most receptive to transfection. In this age range, the blood/testis barrier has not yet formed, and there is a relatively high number of spermatogonia relative to the numbers of other cell types, e.g., spermatids, etc. See J. Kumaran et al., 1949. Poultry Sci., 29:511-520. See also E. Oakberg, 1956. Am. J. Anatomy, 99:507-515; and P. Kluin et al., 1984. Anat. Embryol., 169:73-78.

The transposon-based vectors may be introduced into a testis in an amount ranging from about 0.1  $\mu$ g to 10  $\mu$ g, preferably 1  $\mu$ g to 10  $\mu$ g, more preferably 3  $\mu$ g to 10  $\mu$ g. In a quail, about 5  $\mu$ g is a preferred amount. In a chicken, about 5  $\mu$ g to 10  $\mu$ g per testis is preferred. These amounts of vector DNA may be injected in one dose or multiple doses and at one site or multiple sites in the testis. In a preferred embodiment, the vector DNA is administered at multiple sites in a single testis, both testes being injected in this manner. In one embodiment, injection is spread over three injection sites: one at each end of the testis, and one in the middle. Additionally, the transposon-based vector solution volume may be between approximately 1  $\mu$ l to 75  $\mu$ l in quail and between approximately 1  $\mu$ l to 500  $\mu$ l in chicken. In a preferred embodiment, the transposon-based vector solution volume may be between approximately 20  $\mu$ l to 60  $\mu$ l in quail and between approximately 50  $\mu$ l to 250  $\mu$ l in chicken. Both the amount of vector DNA and the total volume injected into each testis may be determined based upon the age and size of the bird.

10

15

20

25

30

According to the present invention, the polynucleotide cassette is administered in conjunction with an acceptable carrier and/or transfection reagent. Acceptable carriers include, but are not limited to, water, saline, Hanks Balanced Salt Solution (HBSS), Tris-EDTA (TE) and lyotropic liquid crystals. Transfection reagents commonly known to one of ordinary skill in the art that may be employed include, but are not limited to, the following: cationic lipid transfection reagents, cationic lipid mixtures, polyamine reagents, liposomes and combinations thereof; SUPERFECT®, Cytofectene, BioPORTER®, GenePORTER®, NeuroPORTER®, and perfectin from Gene Therapy Systems; lipofectamine, cellfectin, DMRIE-C oligofectamine, and PLUS reagent from InVitrogen; Xtreme gene, fugene, DOSPER and DOTAP from Roche: Lipotaxi and Genejammer from Strategene; and Escort from SIGMA. In one embodiment, the transfection reagent is SUPERFECT®. The ratio of DNA to transfection reagent may vary based upon the method of administration. In one embodiment, a transposon-based vector containing a polynucleotide cassette is administered intratesticularly and the ratio of DNA to transfection reagent can be from 1:1.5 to 1:15, preferably 1:2 to 1:10, all expressed as wt/vol. Transfection may also be accomplished using other means known to one of ordinary skill in the art, including without limitation electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

Depending upon the cell or tissue type targeted for transfection, the form of the transposon-based vector may be important. Plasmids harvested from bacteria are generally closed circular supercoiled molecules, and this is the preferred state of a vector for gene delivery because of the ease of preparation. In some instances, transposase expression and insertion may be more efficient in a relaxed, closed circular configuration or in a linear configuration. In still other instances, a purified transposase protein may be co-injected with a transposon-based vector containing the gene of interest for more immediate insertion. This could be accomplished by using a transfection reagent complexed with both the purified transposase protein and the transposon-based vector.

# Testing for and Breeding Animals Carrying the Transgene

5

10

15

20

25

30

Following administration of a polynucleotide cassette to an animal, DNA is extracted from the animal to confirm integration of the genes of interest. Advantages provided by the present invention include the high rates of integration, or incorporation, and transcription of the gene of interest when administered to a bird via an intraoviduct or intraovary route (including intraarterial administrations to arteries leading to the oviduct or ovary) and contained within a transposon-based vector.

Actual frequencies of integration can be estimated both by comparative strength of the PCR signal, and by histological evaluation of the tissues by quantitative PCR. Another method for estimating the rate of transgene insertion is the so-called primed in situ hybridization technique (PRINS). This method determines not only which cells carry a transgene of interest, but also into which chromosome the gene has inserted, and even what portion of the chromosome. Briefly, labeled primers are annealed to chromosome spreads (affixed to glass slides) through one round of PCR, and the slides are then developed through normal in situ hybridization procedures. This technique combines the best features of in situ PCR and fluorescence in situ hybridization (FISH) to provide distinct chromosome location and copy number of the gene in question. The 28s rRNA gene will be used as a positive control for spermatogonia to confirm that the technique is functioning properly. Using different fluorescent labels for the transgene and the 28s gene causes cells containing a transgene to fluoresce with two different colored tags.

Breeding experiments may also be conducted to determine if germline transmission of the transgene has occurred. In a general bird breeding experiment

performed according to the present invention, each male bird is exposed to 2-3 different adult female birds for 3-4 days each. This procedure is continued with different females for a total period of 6-12 weeks. Eggs are collected daily for up to 14 days after the last exposure to the transgenic male, and each egg is incubated in a standard incubator. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR.

5

10

15

20

25

Any male producing a transgenic embryo is bred to additional females. Eggs from these females are incubated, hatched, and the chicks tested for the exogenous DNA. Any embryos that die are necropsied and examined directly for the transgene or protein encoded by the transgene, either by fluorescence or PCR. The offspring that hatch and are found to be positive for the exogenous DNA are raised to maturity. These birds are bred to produce further generations of transgenic birds, to verify efficiency of the transgenic procedure and the stable incorporation of the transgene into the germ line. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR.

It is to be understood that the above procedure can be modified to suit animals other than birds and that selective breeding techniques may be performed to amplify gene copy numbers and protein output.

# Production of Desired Proteins and Multimeric Proteins in Egg White

In one embodiment, a transposon-based vector containing a polynucleotide cassette of the present invention may be administered to a bird for production of desired proteins or peptides in the egg white. These trassposon-based vectors preferably contain one or more of an ovalbumin promoter, an ovomucoid promoter, an ovalbumin signal sequence and an ovomucoid signal sequence. Oviduct-specific ovalbumin promoters are described in B. O'Malley et al., 1987. EMBO J., vol. 6, pp. 2305-12; A. Qiu et al., 1994. Proc. Nat. Acad. Sci. (USA), vol. 91, pp. 4451-4455; D. Monroe et al., 2000. Biochim. Biophys. Acta, 1517 (1):27-32; H. Park et al., 2000. Biochem., 39:8537-8545; and T. Muramatsu et al., 1996. Poult. Avian Biol. Rev., 6:107-123.

# 30 Production of Desired Proteins and Multimeric Proteins in Egg Yolk

The present invention is particularly advantageous for production of recombinant peptides and proteins of low solubility in the egg yolk. Such proteins include, but are not limited to, membrane-associated or membrane-bound proteins,

lipophilic compounds; attachment factors, receptors, and components of second messenger transduction machinery. Low solubility peptides and proteins are particularly challenging to produce using conventional recombinant protein production techniques (cell and tissue cultures) because they aggregate in water-based, hydrophilic environments. Such aggregation necessitates denaturation and refolding of the recombinantly-produced proteins, which may deleteriously affect their structure and function. Moreover, even highly soluble recombinant peptides and proteins may precipitate and require denaturation and renaturation when produced in sufficiently high amounts in recombinant protein production systems. The present invention provides an advantageous resolution of the problem of protein and peptide solubility during production of large amounts of recombinant proteins.

5

10

15

20

25

30

In one embodiment of the present invention, deposition of a desired protein into the egg yolk is accomplished by attaching a sequence encoding a protein capable of binding to the yolk vitellogenin receptor to a gene of interest that encodes a desired protein. This polynucleotide cassette can be used for the receptor-mediated uptake of the desired protein by the oocytes. In a preferred embodiment, the sequence ensuring the binding to the vitellogenin receptor is a targeting sequence of a vitellogenin protein. The invention encompasses various vitellogenin proteins and their targeting sequences. In a preferred embodiment, a chicken vitellogenin protein targeting sequence is used, however, due to the high degree of conservation among vitellogenin protein sequences and known cross-species reactivity of vitellogenin targeting sequences with their egg-yolk receptors, other vitellogenin targeting sequences can be substituted. One example of a construct for use in the transposon-based vectors of the present invention and for deposition of an insulin protein in an egg yolk is a transposon-based vector containing a vitellogenin promoter, a vitellogenin targeting sequence, a TAG sequence, a pro-insulin sequence and a synthetic polyA sequence. The present invention includes, but is not limited to, vitellogenin targeting sequences residing in the N-terminal domain of vitellogenin, particularly in lipovitellin I. In one embodiment, the vitellogenin targeting sequence contains the polynucleotide sequence of SEQ ID NO:77.

In a preferred embodiment, the transposon-based vector contains a transposase gene operably-linked to a constitutive promoter and a gene of interest operably-linked to a liver-specific promoter and a vitellogenin targeting sequence.

#### Isolation and Purification of Desired Proteins and Multimeric Proteins

5

10

15

20

25

30

For large-scale production of protein, an animal breeding stock that is homozygous for the transgene is preferred. Such homozygous individuals are obtained and identified through, for example, standard animal breeding procedures or PCR protocols.

Once expressed, peptides, polypeptides and proteins can be purified according to standard procedures known to one of ordinary skill in the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis, high performance liquid chromatography, immunoprecipitation and the like. Substantially pure compositions of about 50 to 99% homogeneity are preferred, and 80 to 95% or greater homogeneity are most preferred for use as therapeutic agents.

In one embodiment of the present invention, the animal in which the desired protein is produced is an egg-laying animal. In a preferred embodiment of the present invention, the animal is an avian and a desired peptide, polypeptide or protein is isolated from an egg white. Egg white containing the exogenous protein or peptide is separated from the yolk and other egg constituents on an industrial scale by any of a variety of methods known in the egg industry. See, e.g., W. Stadelman et al. (Eds.), Egg Science & Technology, Haworth Press, Binghamton, NY (1995). Isolation of the exogenous peptide or protein from the other egg white constituents is accomplished by any of a number of polypertide isolation and purification methods well known to one of ordinary skill in the art. These techniques include, for example, chromatographic methods such as gel permeation, ion exchange, affinity separation, metal chelation, HPLC, hydrophobic interaction chromatography, and the like, either alone or in combination. Such techniques are known to one of ordinary skill in the art of protein purification. Another means that may be used for isolation or purification, either in lieu of or in addition to chromatographic separation methods, includes Successful isolation and purification is confirmed by standard electrophoresis. analytic techniques, including HPLC, mass spectroscopy, and spectrophotometry. These separation methods are often facilitated if the first step in the separation is the removal of the endogenous ovalbumin fraction of egg white, as doing so will reduce the total protein content to be further purified by about 50%.

To facilitate or enable purification of a desired protein or peptide, the polynucleotide cassettes may include one or more additional epitopes or domains.

Such epitopes or domains include DNA sequences encoding enzymatic, chemical or photolabile cleavage sites including, but not limited to, an enterokinase cleavage site; the glutathione binding domain from glutathione S-transferase; polylysine; hexahistidine or other cationic amino acids, and sites cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid; thioredoxin; hemagglutinin antigen; maltose binding protein; a fragment of gp41 from HIV; and other purification epitopes or domains commonly known to one of skill in the art. Other proteolytic cleavage sites that may be included in the polynucleotide cassettes are cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Self-splicing cleavage sites such as inteins may also be included in the polynucleotide cassettes of the present invention.

5

10

15

20

25

30

In one representative embodiment, purification of desired proteins from egg white utilizes the antigenicity of the ovalbumin carrier protein and particular attributes of a TAG linker sequence that spans ovalbumin and the desired protein. The TAG sequence is particularly useful in this process because it contains 1) a highly antigenic epitope, a fragment of gp41 from HIV, allowing for stringent affinity purification, and, 2) a recognition site for the protease enterokinase immediately juxtaposed to the desired protein. In a preferred embodiment, the TAG sequence comprises approximately 50 amino acids. A representative TAG sequence is provided below.

The underlined sequences were taken from the hairpin loop domain of HIV gp-41 (SEQ ID NO:87). Sequences in italics represent the cleavage site for enterokinase (SEQ ID NO:89). The spacer sequence upstream of the loop domain was made from repeats of (Pro Ala Asp Asp Ala) (SEQ ID NO:85) to provide free rotation and promote surface availability of the hairpin loop from the ovalbumin carrier protein.

Isolation and purification of a desired protein is performed as follows:

1. Enrichment of the egg white protein fraction containing ovalbumin and the transgenic ovalbumin-TAG-desired protein.

- 2. Size exclusion chromatography to isolate only those proteins within a narrow range of molecular weights (a further enrichment of step 1).
- Ovalbumin affinity chromatography. Highly specific antibodies to ovalbumin will eliminate virtually all extraneous egg white proteins except ovalbumin and the transgenic ovalbumin-TAG-desired protein.

10

15

20

25

30

- 4. gp41 affinity chromatography using anti-gp41 antibodies. Stringent application of this step will result in virtually pure transgenic ovalbumin-TAG-desired protein.
- 5. Cleavage of the transgene product can be accomplished in at least one of two ways:
  - a. The transgenic ovalbumin-TAG-desired protein is left attached to the gp41 affinity resin (beads) from step 4 and the protease enterokinase is added. This liberates the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence is retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), leaves the desired protein together with enterokinase in solution. Enterokinase is recovered and reused.
  - b. Alternatively, enterokinase is immobilized on resin (beads) by the addition of poly-lysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-desired protein eluted from the affinity column of step 4 is then applied to the protease resin. Protease action cleaves the ovalbumin-TAG sequence from the desired protein and leaves both entities in solution. The immobilized enterokinase resin is recharged and reused.
  - c. The choice of these alternatives is made depending upon the size and chemical composition of the transgene target protein.
- 6. A final separation of either of these two (5a or 5b) protein mixtures is made using size exclusion, or enterokinase affinity chromatography. This step allows for desalting, buffer exchange and/or polishing, as needed.

Cleavage of the transgene product (ovalbumin-TAG-desired protein) by enterokinase, then, results in two products: ovalbumin-TAG and the desired protein.

More specific methods for isolation using the TAG label is provided in the Examples. Some desired proteins may require additions or modifications of the above-described approach as known to one of ordinary skill in the art. The method is scaleable from the laboratory bench to pilot and production facility largely because the techniques applied are well documented in each of these settings.

It is believed that a typical chicken egg produced by a transgenic animal of the present invention will contain at least 0.001 mg, from about 0.001 to 1.0 mg, or from about 0.001 to 100.0 mg of exogenous protein, peptide or polypeptide, in addition to the normal constituents of egg white (or possibly replacing a small fraction of the latter). One of skill in the art will recognize that after biological expression or purification, the desired proteins, fragments thereof and peptides may possess a conformation substantially different than the native conformations of the proteins, fragments thereof and peptides. In this case, it is often necessary to denature and reduce protein and then to cause the protein to re-fold into the preferred conformation. Methods of reducing and denaturing proteins and inducing re-folding are well known to those of skill in the art.

## Production of Proteins and Multimeric Proteins in Milk

5

10

15

20

25

30

In addition to methods of producing eggs containing transgenic proteins or peptides, the present invention encompasses methods for the production of milk containing transgenic proteins or peptides. These methods include the administration of a transposon-based vector described above to a mammal, for example through administration into the mammary duct system in order to access epithelial cells. In another embodiment, the transposon-based vector may be administered into arteries supplying the mammary gland. In a further embodiment germline transformed animals may have the transposon-based vector stably incorporated into the mammary In one embodiment, the transposon-based vector contains a epithelial cells. transposase operably-linked to a constitutive promoter and a gene of interest operably-linked to mammary specific promoter. Genes of interest can include, but are not limited to antiviral and antibacterial proteins, immunoglobulins, antibodies and proteins associated with disease and infectious organisms. Milk may be processed for isolation of proteins, multimeric proteins and antibodies using techniques known to one of skill in the art of processing milk and purifying proteins and antibodies from milk.

## Administration of the Vaccines Produced with the Present Invention

The proteins, polypeptides, peptides and antibodies made with the present invention for use in vaccines may be administered to animals or humans using techniques known to one of ordinary skill in the art. Clinical practitioners of ordinary skill in the in the veterinary and human medical fields are familiar with methods, routes, volumes and schedules of vaccine administration, as well as various adjuvants that may be optionally employed to engender or amplify an immune response.

The following examples will serve to further illustrate the present invention without, at the same time, however, constituting any limitation thereof. On the contrary, it is to be clearly understood that resort may be had to various embodiments, modifications and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the invention.

15 EXAMPLE 1

5

10

20

25

30

Preparation of Transposon-Based Vector pTnMod

A vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells, given below as SEQ ID NO:57. The vector of SEQ ID NO:57, termed pTnMod, was constructed and its sequence verified.

This vector employed a cytomegalovirus (CMV) promoter. A modified Kozak sequence (ACCATG) (SEQ ID NO:5) was added to the promoter. The nucleotide in the wobble position in nucleotide triplet codons encoding the first 10 amino acids of transposase was changed to an adenine (A) or thymine (T), which did not alter the amino acid encoded by this codon. Two stop codons were added and a synthetic polyA was used to provide a strong termination sequence. This vector uses a promoter designed to be active soon after entering the cell (without any induction) to increase the likelihood of stable integration. The additional stop codons and synthetic polyA insures proper termination without read through to potential genes downstream.

The first step in constructing this vector was to modify the transposase to have the desired changes. Modifications to the transposase were accomplished with the primers High Efficiency forward primer (Hef) Altered transposase (ATS)-Hef 5' ATCTCGAGACCATGTGTGAACTTGATATTTTACATGATTCTCTTTACC 3'

(SEQ ID NO:91) and Altered transposase- High efficiency reverse primer (Her) 5' GATTGATCATTATCATAATTTCCCCAAAGCGTAACC 3' (SEQ ID NO:92, a reverse complement primer). In the 5' forward primer ATS-Hef, the sequence CTCGAG (SEQ ID NO:93) is the recognition site for the restriction enzyme Xho I, which permits directional cloning of the amplified gene. The sequence ACCATG (SEQ ID NO:5) contains the Kozak sequence and start codon for the transposase and the underlined bases represent changes in the wobble position to an A or T of codons for the first 10 amino acids (without changing the amino acid coded by the codon). Primer ATS-Her (SEQ ID NO:92) contains an additional stop codon TAA in addition to native stop codon TGA and adds a Bcl I restriction site, TGATCA (SEQ ID NO:94), to allow directional cloning. These primers were used in a PCR reaction with pTnLac (p defines plasmid, tn defines transposon, and lac defines the beta fragment of the lactose gene, which contains a multiple cloning site) as the template for the transposase and a FailSafe<sup>TM</sup> PCR System (which includes enzyme, buffers, dNTP's, MgCl<sub>2</sub> and PCR Enhancer; Epicentre Technologies, Madison, WI). Amplified PCR product was electrophoresed on a 1% agarose gel, stained with ethidium bromide, and visualized on an ultraviolet transilluminator. corresponding to the expected size was excised from the gel and purified from the agarose using a Zymo Clean Gel Recovery Kit (Zymo Research, Orange, CA). Purified DNA was digested with restriction enzymes Xho I (5') and Bcl I (3') (New England Biolabs, Beverly, MA) according to the manufacturer's protocol. Digested DNA was purified from restriction enzymes using a Zymo DNA Clean and Concentrator kit (Zymo Research).

10

15

20

25

30

Plasmid gWhiz (Gene Therapy Systems, San Diego, CA) was digested with restriction enzymes Sal I and BamH I (New England Biolabs), which are compatible with Xho I and Bcl I, but destroy the restriction sites. Digested gWhiz was separated on an agarose gel, the desired band excised and purified as described above. Cutting the vector in this manner facilitated directional cloning of the modified transposase (mATS) between the CMV promoter and synthetic polyA.

To insert the mATS between the CMV promoter and synthetic polyA in gWhiz, a Stratagene T4 Ligase Kit (Stratagene, Inc. La Jolla, CA) was used and the ligation set up according to the manufacturer's protocol. Ligated product was transformed into *E. coli* Top10 competent cells (Invitrogen Life Technologies,

5

10

15

20

25

30

Carlsbad, CA) using chemical transformation according to Invitrogen's protocol. Transformed bacteria were incubated in 1 ml of SOC (GIBCO BRL, CAT# 15544-042) medium for 1 hour at 37° C before being spread to LB (Luria-Bertani media (broth or agar)) plates supplemented with 100 µg/ml ampicillin (LB/amp plates). These plates were incubated overnight at 37° C and resulting colonies picked to LB/amp broth for overnight growth at 37° C. Plasmid DNA was isolated using a modified alkaline lysis protocol (Sambrook et al., 1989), electrophoresed on a 1% agarose gel, and visualized on a U.V. transilluminator after ethidium bromide staining. Colonies producing a plasmid of the expected size (approximately 6.4 kbp) were cultured in at least 250 ml of LB/amp broth and plasmid DNA harvested using a Qiagen Maxi-Prep Kit (column purification) according to the manufacturer's protocol (Oiagen, Inc., Chatsworth, CA). Column purified DNA was used as template for sequencing to verify the changes made in the transposase were the desired changes and no further changes or mutations occurred due to PCR amplification. sequencing, Perkin-Elmer's Big Dye Sequencing Kit was used. All samples were sent to the Gene Probes and Expression Laboratory (LSU School of Veterinary Medicine) for sequencing on a Perkin-Elmer Model 377 Automated Sequencer.

Once a clone was identified that contained the desired mATS in the correct orientation, primers CMVf-NgoM IV (5' TTGCCGGCATCAGATTGGCTAT (SEQ ID NO:95); underlined bases denote NgoM IV recognition site) and Syn-polyA-BstE II (5' AGAGGTCACCGGGTCAATTCTTCAGCACCTGGTA (SEQ ID NO:96); underlined bases denote BstE II recognition site) were used to PCR amplify the entire CMV promoter, mATS, and synthetic polyA for cloning upstream of the transposon in pTnLac. The PCR was conducted with FailSafe<sup>TM</sup> as described above, purified using the Zymo Clean and Concentrator kit, the ends digested with NgoM IV and BstE II (New England Biolabs), purified with the Zymo kit again and cloned upstream of the transposon in pTnLac as described below.

Plasmid pTnLac was digested with NgoM IV and BstE II to remove the ptac promoter and transposase and the fragments separated on an agarose gel. The band corresponding to the vector and transposon was excised, purified from the agarose, and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs) to prevent self-annealing. The enzyme was removed from the vector using a Zymo DNA Clean and Concentrator-5. The purified vector and CMVp/mATS/polyA

were ligated together using a Stratagene T4 Ligase Kit and transformed into E. coli as described above.

Colonies resulting from this transformation were screened (mini-preps) as describe above and clones that were the correct size were verified by DNA sequence analysis as described above. The vector was given the name pTnMod (SEQ ID NO:57) and includes the following components:

5

10

15

20

25

30

Base pairs 1-130 are a remainder of F1(-) on from pBluescriptll sk(-) (Stratagene), corresponding to base pairs 1-130 of pBluescriptll sk(-).

Base pairs 131 - 132 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 133 -1777 are the CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems), corresponding to bp 229-1873 of pGWiz. The CMV promoter was modified by the addition of an ACC sequence upstream of ATG.

Base pairs 1778-1779 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 1780 - 2987 are the coding sequence for the transposase, modified from Tn10 (GenBank accession J01829) by optimizing codons for stability of the transposase mRNA and for the expression of protein. More specifically, in each of the codons for the first ten amino acids of the transposase, G or C was changed to A or T when such a substitution would not alter the amino acid that was encoded.

Base pairs 2988-2993 are two engineered stop codons.

Base pair 2994 is a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 2995 - 3410 are a synthetic polyA sequence taken from the pGWiz vector (Gene Therapy Systems), corresponding to bp 1922-2337 of 10 pGWiz.

Base pairs 3415 - 3718 are non-coding DNA that is residual from vector pNK2859.

Base pairs 3719 - 3761 are non-coding  $\lambda$  DNA that is residual from pNK2859.

Base pairs 3762 - 3831 are the 70 bp of the left insertion sequence recognized by the transposon Tn10.

Base pairs 3832-3837 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 3838 - 4527 are the multiple cloning site from pBluescriptll sk(20), corresponding to bp 924-235 of pBluescriptll sk(-). This multiple cloning site may be used to insert any coding sequence of interest into the vector.

Base pairs 4528-4532 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 4533 - 4602 are the 70 bp of the right insertion sequence recognized by the transposon Tn10.

Base pairs 4603 - 4644 are non-coding  $\lambda$  DNA that is residual from pNK2859.

Base pairs 4645 - 5488 are non-coding DNA that is residual from pNK2859.

Base pairs 5489 - 7689 are from the pBluescriptll sk(-) base vector - (Stratagene, Inc.), corresponding to bp 761-2961 of pBluescriptll sk(-).

Completing pTnMod is a pBlueScript backbone that contains a colE I origin of replication and an antibiotic resistance marker (ampicillin).

It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

All plasmid DNA was isolated by standard procedures. Briefly, *Escherichia coli* containing the plasmid was grown in 500 mL aliquots of LB broth (supplemented with an appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA was recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc., Chatsworth, CA) according to the manufacturer's protocol. Plasmid DNA was resuspended in 500  $\mu$ L of PCR-grade water and stored at -20°C until used.

#### **EXAMPLE 2**

### 25 Transposon-Based Vector pTnMCS

5

10

15

20

30

Another transposon-based vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells. This vector was termed pTnMCS and its constituents are provided below. The sequence of the pTnMCS vector is provided in SEQ ID NO:56. The pTnMCS vector contains an avian optimized polyA sequence operably-linked to the transposase gene. The avian optimized polyA sequence contains approximately 75 nucleotides that precede the A nucleotide string.

Bp 1 - 130 Remainder of F1 (-) ori of pBluescriptII sk(-) (Stratagene) bp1-130

Bp 133 – 1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems) bp 229-1873

Bp 1783 - 2991 Transposase, from Tn10 (GenBank accession #J01829) bp 108-1316

Bp 2992 - 3344 Non coding DNA from vector pNK2859

5 Bp 3345 – 3387 Lambda DNA from pNK2859

Bp 3388 – 3457 70 bp of IS10 left from Tn10

Bp 3464 - 3670 Multiple cloning site from pBluescriptII sk(-), thru the XmaI site bp 924-718

Bp 3671 - 3715 Multiple cloning site from pBluescriptII sk(-), from the XmaI site thru the XhoI site. These base pairs are usually lost when cloning into pTnMCS bp 717-673

Bp 3716 – 4153 Multiple cloning site from pBluescriptII sk(-), from the XhoI site bp 672-235

Bp 4159 - 4228 70 bp of IS10 right from Tn10

15 Bp 4229 - 4270 Lambda DNA from pNK2859

Bp 4271 – 5114 Non-coding DNA from pNK2859

Bp 5115 - 7315 pBluescript sk (-) base vector (Stratagene, Inc.) bp 761-2961.

#### **EXAMPLE 3**

20 Production of Antibody in Egg White

10

25

30

A transposon-based vector containing a CMV promoter/cecropin prepro/antibody heavy chain/cecropin pro/Antibody light chain/conalbumin poly A (SEQ ID NO:97) was injected into the oviduct of quail and chickens. A total of 20 birds were injected (10 chickens and 10 quail) and eggs were harvested from the birds once the eggs were laid. Partially purified egg white protein (EW) was then run on a gel under both reducing and non-reducing conditions. Figure 5 is a picture of the gel. Lanes 1 & 18: molecular weight markers, Lanes 2 and 3: EW #1, non-reduced, reduced, respectively; Lanes 4 and 5: EW #2, non-reduced, reduced, respectively, Lanes 8 and 9: EW #4, non-reduced, reduced, respectively; Lanes 10 and 11: EW #5, non-reduced, reduced, respectively; Lanes 12 and 13: EW #6, non-reduced, reduced, respectively; Lanes 14 and 15: EW #7, non-reduced, reduced, respectively; and Lanes 16 and 17: EW #8 Control, non-reduced, reduced, respectively. Based upon the gel results, the

possibility that the egg white in the treated chicken and quail contains antibody produced by the above-mentioned transposon-based vector cannot be excluded.

#### **EXAMPLE 4**

5 Additional Transposon-Based Vectors for Administration to an Animal

The following example provides a description of various transposon-based vectors of the present invention and several constructs for insertion into the transposon-based vectors of the present invention, all for intraoviduct administration. These examples are not meant to be limiting in any way. The constructs for insertion into a transposon-based vector are provided in a cloning vector pTnMCS or pTnMod, both described above.

# pTnMOD (CMV-prepro-HCPro-Lys-CPA) (SEQ ID NO:97)

Bp 1-4090 from vector pTnMod, bp 1 - 4090

10

15

20

30

Bp 4096-5739 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5746-5916 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733
Bp 5923-7287 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7288-7302 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 7309-7953 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7960-8372 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

25 Bp 8374-11973 from cloning vector pTnMod, bp 4091-7690

# pTnMCS (CHOVep-prepro-HCPro-CPA) (SEQ ID NO:98)

Bp 1-3715 from vector pTnMCS, bp 1-3715

Bp 3721-4395 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675

Bp 4402-5738 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 5745-5915 Capsite/Prepro taken fron GenBank accession # X07404, bp 563-733

Bp 5922-7286 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

- Bp 7287-7298 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)
- 5 Bp 7305-7949 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
  - Bp 7956-8363 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
  - Bp 8365-11964 from cloning vector pTnMCS, bp 3716-7315

10

## pTnMCS(CHOvep-prepro-HCPro-Lys-CPA) (SEQ ID NO:99)

- Bp 1 3715 from vector pTnMCS, bp 1-3715
- Bp 3721 4395 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675
- 15 Bp 4402 5738 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336
  - Bp 5745 5915 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733 Bp 5922 7286 Heavy Chain gene construct taken from antibody RM2 provided by
  - Mark Glassy (Shantha West, Inc)
- 20 Bp 7287 7301 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)
  - Bp 7308 7952 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
  - Bp 7959 8366 Conalbumin polyA taken from GenBank accession # Y00407, bp
- 25 10651-11058
  - Bp 8368 11967 from cloning vector pTnMCS, bp 3716-7315

### pTnMCS (CMV-prepro-HCPro-CPA) (SEQ ID NO:100)

- Bp 1-3715 from vector pTnMCS, bp 1-3715
- 30 Bp 3721 5364 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
  - Bp 5371-5541 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

Bp 5548 - 6912 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

- Bp 6913 6924 Pro taken from GenBank accession # X07404, bp 719-730 (does not Lysine)
- 5 Bp 6931 7575 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
  - Bp 7582 7989 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
  - Bp 7991 11590 from cloning vector pTnMCS, bp 3716-7315

10

# pTnMCS (CMV-prepro-HCPro-Lys-CPA) (SEQ ID NO:101)

- Bp 1 3715 from vector pTnMCS, bp 1-3715
- Bp 3721 5364 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
- Bp 5371-5541 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733
  Bp 5548 6912 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
  - Bp 6913 6927 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)
- 20 Bp 6934 7578 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
  - Bp 7585 7992 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
  - Bp 7994 11593 from cloning vector pTnMCS, bp 3716-7315

25

## pTnMod (CHOvep-prepro-HCPro-CPA) (SEQ ID NO:102)

- Bp 1-4090 from vector pTnMod, bp 1-4090
- Bp 4096-4770 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675
- 30 Bp 4777-6113 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336
  - Bp 6120-6290 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

Bp 6297-7661 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7662-7673 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)

5 Bp 7680-8324 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 8331-8738 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 8740-12339 from cloning vector pTnMod, bp 3716-7315

10

## pTnMod (CHOvep-prepro-HCPro-LYS-CPA) (SEQ ID NO:103)

Bp 1-4090 from vector pTnMod, bp 1-4090

Bp 4096-4770 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675

Bp 4777-6113 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 6120-6290 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733 Bp 6297-7661 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7662-7676 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 7683-8327 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 8334-8741 Conalbumin polyA taken from GenBank accession # Y00407, bp

25 10651-11058

Bp 8743-12342 from cloning vector pTnMod, bp 3716-7315

#### pTnMod (CMV-prepro-HCPro-CPA) (SEQ ID NO:104)

Bp 1–4090 from vector pTnMod, bp 1-4090

30 Bp 4096-5739 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5746-5916 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

Bp 5923-7287 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7288-7299 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)

5 Bp 7306-7950 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7557-7969 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 7971-11970 from cloning vector pTnMod, bp 3716-7315

10

15

20

25

30

### Example 5

Production of a Vaccine in Transgenic Chickens against Brucella Melitensis, Collection of the Vaccine from Eggs and Use of the Vaccine to Confer a Protective Response

Brucella species are considered potential biological warfare agents due to the characteristic protracted debilitating disease caused by these organisms. The antigens of interest are the Group 3 Omps which have been shown to demonstrate a role in protective immunity. Group 3 Omps are comprised of two separate proteins of 25 kDa (Omp25) and 31 kDa (Omp31) (Cloeckaert et al., 1996a).

First, genes encoding *Brucella melitensis* outer membrane proteins (Omp 25 and Omp 31) were cloned for expression into egg albumin. The vectors that are used have been described above and the genes of interest encode for Omp 25 and Omp 31.

Harvest of Brucella Genomic DNA: Genomic DNA from B. melitensis is harvested from bacteria grown in 250 ml of Brucella Broth (Difco Laboratories, Detroit, MI) for 18 hours at 37° C. The culture is killed by the addition of chloroform for 1 h to prevent accidental infection of lab personnel. After killing, genomic DNA isharvested using a Qiagen Genomic-Tip System (Qiagen, Inc., Valencia, CA) according to the manufacturer's protocol and resuspended in PCR grade water.

PCR Optimization: For PCR optimization, primers specific to Omp25 and Omp31 are used to amplify each gene from genomic DNA of *B. melitensis*. Conditions for PCR with each set of primers are optimized using a FailSafe<sup>TM</sup> PCR System kit (Epicentre Technologies, Madison, WI), the PCR fragments separated by agarose gel electrophoresis, excised from the gel, and purified from the agarose using

a QIAquick Gel Extraction Kit (Qiagen, Inc). Purified DNA corresponding to Omp25 and Omp31 is stored at -20° C until needed.

Plasmid DNA Isolation: All plasmid DNA is isolated by procedures routinely used in the laboratory of Dr. Cooper. Briefly, *Escherichia coli* containing the desired plasmid is grown in 500 ml aliquots of LB broth (supplemented with the appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA is recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc.) according to the manufacturer's protocol. Plasmid DNA is resuspended in 500  $\mu$ L of PCR grade water and stored at  $^{-}$  20°C until needed.

5

10

15

20

25

30

The vectors are digested with restriction enzymes. Bands corresponding to the vector are excised from the gel, purified from agarose (see above), and ligated (Fast-Link<sup>TM</sup> DNA Ligation Kit, Epicentre Technologies) overnight with either DNA encoding Omp25 or Omp31. Ligation mixtures are electroporated into  $E.\ coli$  STBL4<sup>TM</sup> (Life Technologies, Rockville, MD), plated onto LB agar plates supplemented with 200  $\mu$ g/ml ampicillin, and incubated at 37° C overnight.

Resulting colonies are grown at 37° C overnight in 2 ml of LB broth supplemented with 200 µg/ml ampicillin, plasmid DNA harvested using a QIAprep Spin Miniprep Kit (Qiagen, Inc.), and separated by agarose gel electrophoresis. Bacterial isolates containing the expected insert size are frozen at  $^-80^{\circ}$  C. The insert is sequenced to determine orientation and verify the sequence. All sequencing is conducted using a BigDye<sup>TM</sup> Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA) according to the manufacturer's protocol and loaded onto a PE Applied Biosystems 310 Genetic Analyzer. Isolates containing Omp25 (pTnOv25) and Omp31 (pTnOv31) with the correct sequence and orientation are grown as described above for Maxi-Prep DNA isolation. Purified DNA will be resuspended in TE buffer (10 mM tris, 5 mM EDTA, pH 8.0) and stored at  $^-20^{\circ}$  C until needed.

Next transgenic chickens are developed to produce Omp25 and Omp31 in the albumin portion of their eggs. The experimental males are obtained from commercial sources at one day of age and maintained in brooders until used. After being placed in a treatment group, males will be housed individually in standard caging in temperature controlled spaces appropriate to the age of the birds. They are watered and fed standard Starter Chick chow ad lib and kept initially in a 23:1 h light/dark

cycle, stepping down at approximately weekly intervals to a 15h:8h light/dark cycle, as this regimen optimizes sexual maturity and fertility (Duncan 1999).

For transfection, plasmid DNA is complexed with Superfect<sup>TM</sup> (Qiagen, Inc) in an approximate 1:3 ratio, following the manufacturer's protocol. The DNA is injected (described below) at a rate of 5  $\mu$ g per testis in a 50  $\mu$ L volume [Hank's balanced salt solution (HBSS) will be used to bring the total volume to 50  $\mu$ l]. At least 20 males are used for controls and receive an injection of HBSS and transfecting reagent. All birds are held until sexual maturity before sperm samples are obtained for PCR.

5

10

15

20

25

30

At age 30-98 days, groups of individual males are subjected to transgene delivery by direct intratesticular injection of DNA. Males are anesthetized with isoflourane via a gas machine and positioned on their side on an animal board with cords tractioning wings and feet to allow access to the testes area. The area is swabbed with 0.5% chlorhexidine and a 2 cm dorsolateral incision made over the testis (which is identical to the procedure for caponization). After exposing the testis, a capillary injection device (described below) is used to inject, in 2 or 3 sites, about 50 µl of the DNA/SuperFect complex (Qiagen) (~5 µg DNA per testis). The injections are calculated to suffuse the DNA throughout the whole testis, allowing as much contact of DNA with spermatogonia as possible. Following injection, the wound is closed in two layers with 4-6 absorbable suture, and the contralateral testis similarly exposed and injected. Following surgery, the bird is returned to its cage to recover. Approximately 50 males are used for this experimental regimen to increase the likelihood of obtaining a male having transgenic sperm.

The injection device consists of a 27 ga hypodermic needle cut off at the hub, bent to the desired angle, and welded onto a standard 100 µl capillary tube. The capillary is filled with the DNA solution and placed into a standard mouth pipette so that the injection rate and volume can be accurately controlled hands-free. After injection, males are allowed to reach sexual maturity in order to obtain sperm for PCR. Males with sperm positive for Omp25 or Omp31 are mated with females to produce transgenic F1s. The F1s are verified for copy number of the transgene. These transgenic F1s are raised to maturity.

Next the expressed proteins are purified from the eggs of the transgenic chickens and subsequently used in vaccine efficacy trials. The purification

is achieved with methods presented above for purification of proteins from eggs. The following is only one approach to purify the protein. The initial objective in the purification of expressed proteins in egg white is to eliminate the major component of egg white, namely, the albumin proteins. Li — Chan, et al., 1986 have described a versatile and simple ion — exchange procedure which largely accomplishes this elimination of the albumin.

5

10

15

20

25

30

The egg white is lightly homogenized to reduce viscosity [shear rate of 3140 sec<sup>-1</sup> reduces albumin viscosity from 21 to 5 centipoises (Tung et al., 1969]. The albumin solution is filtered through two layers of muslin to remove chalazae. The resulting solution is then passed through an industrial cation – exchanger, IMAC-HP333 (Rohm and Haas, Philadelphia, PA) which has been equilibrated at pH 9 with low ionic strength buffer (phosphate or bicarbonate). The albumin passes through the ion – exchange column leaving some ovomucin and all of the avidin and lysozyme on the column.

This procedure allows cationic proteins/peptides to be separated from the albumin at an early stage in the purification. What is required for making this process a general one is to either purify such proteins which are by their nature, highly positively charged (e.g. cecropin – like lytic peptides or their derivatives) or utilize fusion proteins wherein the fusion adduct contains the requisite cationic charge. Also required is that the magnitule of charge on the expressed protein or fusion protein must exceed considerably that of any possible cationic contaminant.

Once the cationic entities have been captured by the ion – exchanger, elution is initiated by a series of increasing ionic strength, pH 8.0 bicarbonate buffers. The desired product is retained until it is selectively eluted at the highest ionic strength (ammonium bicarbonate, approaching 2M). The water and most of the ammonium bicarbonate are removed by lyophilization. The recovered fusion protein is dissolved in 88% formic acid and subjected to cyanogen bromide cleavage. If the desired component is anionic, the lyophilized residue is dissolved in dilute acetic acid and passed through and anion – exchange resin contained in a jacketed column. The column is washed with dilute hydrochloric acid (0.05 M) and cold (-5°C) ethylene glycol solution pumped through the jacket on the column. By-products from the cleavage reaction are substantially removed by this step. The anionic protein is then lyophilized (using solid KOH in the vacuum system to capture the HCl) to dryness,

5

10

25

30

dissolved in a minimal amount of water and lyophilized (twice) to remove residual HCl.

If the desired component of the fusion protein is cationic then the cleavage mixture is diluted 5 – fold with acetate buffer and passed onto a sulfoethylcellulose cation exchanger contained in a jacketed column. The column is washed with 0.05 M HCl and then 0.15 M HCl to remove residual cationic materials including degradation products while pumping cold (-5°C) ethylene glycol through the column jacket. The desired product is eluted with 1.0 M HCL adjusted carefully to pH 2.7 by the extremely slow addition of very cold ammonium hydroxide. The eluted protein is lyophilized to dryness using solid KOH within the vacuum system to capture the HCl. The protein is dissolved in a minimum amount of water and lyophilized (twice) to remove residual HCl. Protein quantity will be determined using a Bio-Rad Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA). Purity is determined with Western blots, HPLC, ELISAs and other methods known to one of ordinary skill in the art.

Vaccination and Challenge of Mice: The BALB/c murine model of brucellosis established by Montaraz and Winter (1986) has provided much of the current understanding of immunity to brucellosis. The model is valuable since the murine colonization profile of virulent type strains and vaccine candidates mimics what is often attenuated in the ruminant host (Montaraz and Winter, 1986; Winter, 1990; Schurig et al., 1991; Elzer et al., 1996; Stevens et al., 1997; Edmonds et al., 2000a; Edmonds et al., 2000b). The murine model is also useful due to practical concerns including availability of reagents, cost, and population homogeneity (Montaraz and Winter, 1986; Garcia-Carrillo, 1990).

Bacterial Strains: *Brucella melitensis* strain 16M has been documented to cause chronic infections in mice (Jiminez et al., 1994; Philips et al., 1995). Infectious doses containing approximately 1 x 10<sup>9</sup> and 1 x 10<sup>11</sup> CFU/ml are prepared as described (Elzer et al., 1994c) and snap frozen in liquid nitrogen prior to storage at -80°C. Immediately prior to animal inoculation, samples are thawed and diluted in sterile saline to the appropriate concentration. Doses are verified by serial dilution and plating on Schaedler agar containing 5% bovine blood (SBA) (Difco Laboratories, Detroit, MI).

Female BALB/c mice at approximately 6 to 8 weeks of age are obtained from the Division of Laboratory Animal Medicine at the School of Veterinary Medicine at Louisiana State University (Baton Rouge, LA) and held at least one week prior to use. All animals are housed in a restricted access facility. All mice are used in accordance with the Animal Care and Use Committee at Louisiana State University, and approved protocols will be followed.

5

10

15

20

25

30

Murine Vaccine Efficacy Trials: Mice are divided into groups of ten and are given the following vaccines intraperitoneal (i.p.): Group 1 – saline control; Group 2 –  $5 \times 10^4$  CFU of B. melitensis vaccine strain Rev-1 i.v.; Group 3 – Titremax adjuvant; Group 4 – Omp25 plus Titremax adjuvant; Group 5 – Omp31 plus Titremax adjuvant; and Group 6 – Omp25 and Omp31 plus Titremax adjuvant. Mice will be vaccinated with  $30 \mu g B$ . melitensis Omp25 or Omp31. The dose of the Rev-1 strain is based on previously published work in BALB/c mice (Jimenez de Bagues et al., 1994a; Jimenez de Bagues et al., 1994b). Prior to vaccination and 8 weeks post-vaccination, mice were bled and the serum pooled for each group and the Omp25 and Omp31 vaccinated mice are boosted.

Ten weeks following initial vaccination with either Omp25, Omp31, Omp25 plus Omp31, or *B. melitensis* strain Rev. 1, both vaccinated and non-vaccinated mice are i.v. inoculated with the challenge strain 5 x 10<sup>4</sup> CFU of *B. melitensis* strain 16M (Jimenez de Bagues et al., 1994a; Jimenez de Bagues et al., 1994b). Two weeks post-challenge, all mice are euthanized by halothane overdose. At necropsy, the whole spleen is removed by aseptic technique for bacteriological analysis while blood is obtained for serological analysis. Spleens are homogenized with a Sorvall Omni-Mixer (Newton, CT) in sterile phosphate-buffered-saline (PBS) and subsequently serially diluted and plated on SBA. All plates are incubated at 37°C in a 5% CO<sub>2</sub> atmosphere for 3 to 4 days until isolated colonies appear.

Serological procedures: Blood is collected from all the mice before inoculation with *B. melitensis* and at two-week intervals thereafter until they are killed. The sera obtained is stored at <sup>80</sup>°C. *Brucella*-specific antibodies are detected by the Card test and Western blot analysis with whole-cell lysates of *B. melitensis* (Elzer et al., 1994a).

Vaccination and challenge of Goats: Strains

5

15

20

25

30

B. melitensis is of established virulence in pregnant goats (Meador and Deyoe 1986, Enright 1990). The construction and phenotypic characteristics of the isogenic 2308 htrA mutant, PHE1, have been described by Elzer et al., (1994a). Infectious doses containing approximately 10<sup>9</sup> or 10<sup>11</sup> colony forming units (cfu) ml<sup>-1</sup> in Brucella broth (Difco, Detroit, MI, USA) were prepared as described by Elzer et al., (1994a) and snap frozen in a dry ice-ethanol bath before being stored at -80°C. These preparations were thawed just before use and their content of viable brucellae was assessed by serial dilution and plating.

Serological procedures: Blood is collected from all the goats before they are vaccinated, and at two-week intervals thereafter until they are killed. The sera obtained is stored at -80°C. *Brucella*-specific antibodies are detected by the Card test and Western blot analysis with whole-cell lysates of *B abortus* (Elzer et al., 1994a).

Vaccine Efficacy Study: Seventy-two mature, non-pregnant female goats are divided into 6 groups and given 100  $\mu$ g of the following vaccines subcutaneously (s.c.): Group 1 – saline control; Group  $2-5 \times 10^4$  CFU of B. melitensis vaccine strain Rev-1 i.v.(this group receives saline first and then Rev-1 2 weeks later); Group 3 - Titremax adjuvant; Group 4 - Omp25 rlus Titremax adjuvant; Group 5 - Omp31 plus Titremax adjuvant: and Group 6 - Omp25 and Omp31 plus Titremax adjuvant. Eight weeks later, the 3 Omp groups and adjuvant control group are boosted with a second dose of the appropriate formulation. Twelve other animals are vaccinated SC with 1 x 10<sup>6</sup> CFU of the vaccine B. melitensis strain Rev. 1. Previous experiments in goats have shown that this dose of strain Rev. 1 provides protection against challenge with virulent B. melitensis strain 16M (Elzer et al., 1998). The remaining 12 animals serve as controls and receive another SC inoculation of the saline diluent. Four weeks after vaccination, a fertile male goat is housed with each group. Following ultrasound confirmation of pregnancy, all animals are conjunctively challenged at approximately 110 days gestation with 1 x 10<sup>7</sup> CFU of virulent B. melitensis strain 16M. Following inoculation, animals are monitored daily for delivery. At parturition, the birth status of the kid is recorded and within 24 h, necropsy samples obtained for bacteriologic analysis from the lung and abomasal fluid. Within one month of delivery, dams are

euthanized and samples taken for bacteriologic analysis from the parotid lymph node, prescapular lymph node, supramammary lymph node, internal iliac lymph node, liver, spleen, milk, and uterine fluid. Tissue samples are frozen at -4°C for later analysis; milk and uterine fluid are swabbed the day of collection. Serum samples are obtained prior to vaccination, at challenge, and at necropsy for serological analysis.

5

10

25

30

Bacteriological procedures: All the tissue specimens are homogenized in sterile PBS (0·002M NaH<sub>2</sub>PO<sub>4</sub>, 0·008M Na<sub>2</sub>HPO<sub>4</sub>, 0·15M NaCl, pH 7·2), plated on to Farrell's selective medium (Farrell 1974) and Schaedler agar supplemented with 5% sheep blood. The plates are incubated for 14 days at 37°C in an atmosphere of 5% carbon dioxide. The isolates were identified as *Brucella* species by their Gram-staining characteristics, colonial morphology, urease and oxidase reactions and agglutination with *B. abortus*-specific antibodies in a standard slide agglutination assay. (Elzer et al., 1994a)

Mice Studies: Mice vaccinated with the subcellular components (OMP) in conjunction with the adjuvant Titermax exhibit increased protective immunity (i.e., lower CFU/gm of tissue) as compared to the saline controls. Serological analysis remains negative since the *Brucella* diagnostic tests are dependent on antibody reactions to the O-side chain of the organism. The adjuvant alone does not contribute to the immune response or interfere with the serological analyses. The combination of the Titermax and OMP vaccines results in a predominant TH1 protective immune response as demonstrated by a lower bacterial load in cultured tissues.

Goat Studies: Goats receiving the Titermax and OMP combinations exhibit a protective immune response post-challenge as demonstrated by decreased colonization of the dam/fetus pairs and/or decreased abortions as compared to the saline or Titermax alone controls. *B. melitensis* Rev 1 is the accepted vaccine for caprine brucellosis; it is the "gold standard" for protection in this model. The Titermax/OMP vaccinations results in equivalent or better protection as compared to Rev 1. However, unlike Rev 1, the Titermax/OMP groups remain serologically negative on all diagnostic tests until after challenge when all groups become positive.

All patents, publications and abstracts cited above are incorporated herein by reference in their entirety. It should be understood that the foregoing relates only to preferred embodiments of the present invention and that numerous modifications or alterations may be made therein without departing from the spirit and the scope of the present invention as defined in the following claims.

## Appendix

SEQ ID NO:3 (cecropin prepro)
AAT TTC TCA AGG ATA TTT TTC TTC GTG TTC GCT
GTT TCG GCT GCG CCA GAG CCG AAA ID NO:2 (cecropin pro extended) CCA GAG CCG AAA TGG AAA GTC TTC AAG SEQ ID NO:1 (cecropin pro) CCA GAG CCG AAA පුදුල SEQ 10

5

TTG GTT CTG GCT TTG TCA ACA

ID NO:4 (cecropin prepro extended)
TTC TCA AGG ATA TTT TTC TTC GTG TTC GCT TTG GTT CTG GCT TTG TCA ACA
TCG GCT GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG SEQ AAT GTT 15

SEQ ID NO:5(modified Kozak sequence) 20

atg ctg ggc atc tgg acc ctc cta cct ctg gtt ctt acg tct gtt gct aga tta SEQ ID NO:6 SEQ ID NO:7 25

atg att cct gcc aga ttt gcc ggg gtg ctg ctt gct ctg gcc ctc att ttg cca ggg acc ctt tgt

gcagagcaat ggtggccagg ctgggggttgg ggctgctgct gctggcactg ctcctaccca cgcagattta ttcc SEQ ID NO:8 atgg 30

atgaatctat cgaacatttc tgcggtaaaa dtattaacac tggtggttag cgctgccatc gct SEQ ID NO:9 35

	OF ON GT ON				
2	atgaccatec tttteettae tatggttate teataettea gttgeatgaa agetgeeeeg atgacatec ttggacceeg atgacatege agetgeeeeg atgacatec teataages agettggett acceaggtet teggacceae gggactettg aaageetaae tgggeeeaat getggtteaa gaggactgae ateaetggeg gacaettttg aacaegtgat agaggagett etagatgaag ateaggacat ceageecagt	ttttccttac tatggttatc ctagtgtaag aggacatggc aaagcctaac tgggcccaat aacacgtgat agaggagctt	agettggett acceaggtet teggaceeg getggtteaa gaggactgae ateactggeg etagatgaag ateaggaeat ceageeeagt	tcatacttca gttgcatgaa agctgccccg agcttggctt acccaggtct tcggacccac gctggttcaa gaggactgac atcactggcg ctagatgaag atcaggacat ccagcccagt	agctgccccg tcggacccac atcactggcg ccagcccagt tcaaqtqcct
10	yayyaaaacc ayyacycy ttggaacccc cactgctc atgtccatga gagtccgg	cactgotott totgotogag gagtooggog toactotgac		gagtacaaaa actacctgga tgctgcaaac	tgctgcaaac
	SEQ ID NO:11 cgtcgctttc gctcttttct cgtcggcgat tgctttctcc	tt tctcgctttc	gotottttct	cgtcggcgat	tgetttetee
15	gacgacgatc cgttgatccg acaa gccgagcatc acttttcact tttt	cgttgatccg acaagttgta tcgggaaacg atgacaacca tatgttaaac acttttcact tttt	tcgggaaacg	atgacaacca	tatgttaaac
	SEQ ID NO:12 atg tccatcttgt tttatgtgat atttcttgca tatcttcgtg gcattcagtc aactaatatg gatcaaagga gtttgccaga agattcaatg aattctctca ttattaaact cattcgggca	tgtgat atttcti ga agattcaatg	igca tatette aattetetea	gtg gcattca ttattaaact	igtc aactaatatg cattcgggca
70	gacatotiga aaaacaag atagigcaga aagiagag ttocaaccaa ttatites	aaaacaagct ttctaagcag gtgatggatg tcaaggaaaa ctatcaaaac aagtagagga ccaccaggag atggatggag atgaaaatgt gaaatcagac ttatttcaat qqatacaqac ctcctaaggc agcagagacg ctacaactct	gtgatggatg atggatggag ctcctaaggc	gtgatggatg tcaaggaaaa atggatggag atgaaaatgt ctcctaaggc agcagagacg	ctatcaaaac gaaatcagac ctacaactct
i d		tectaagtga caacacaca etggaaccac caccactgta gaagtteagt ggtattaaac agaacetete gaaggaaaag	caacacaca ctggaaccac caccactgta ggtattaaac agaacctctc gaaggaaaag	caccactgta gaaggaaaag	cctcacagag gt
23	SEQ ID NO:13	1	\$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$	1000	משמלהמשמ
	atggtgcatc tgactcctga ggagagicci yccyllaciy coolylysyyy curysyyya curysyyya curysyyyy curysyyya chasyyyyy chasy chasy chasy curysyyyy chasy chast chasy chasy chasy chast chas	iga ggagaagici qa qqccctqqqc	aggetgetgg	tggtctaccc	ttggacccag

m

caaccctaag tcacctggac gcacgtggat tcactttggc tgtggctaat tggtggctgg tgtggctaat tccaatttct attaaaggtt gcctttagtg atggcctggc gagctgcact gtgacaagct ctgttatggg tgctggccca aggcrgcryy actcctgatg gggcaacgtg ctggtctgtg gcaggctgcc tatcagaaag ggatctgtcc a gtggatgaag ttggtggtga ggccctggggc aggttctttg agtcctttgg ggatctgtcc cacactgagt gcacctttgc atggcaagaa tcaggctcct aaagaattca ccccaccagt gtgaaggctc cctgagaact aacctcaagg 30

tttcttgctg

ctaagctcgc

gccctggccc acaagtatca

cctttgttcc ctaagtccaa ctactaaact gggggatatt atgaagggcc ttgagcatct ggattctgcc taataaaaaa catttatttt cattgc

SEQ ID NO:14

5

gcatgggggac ggcgcttctc cagcgcgggg gctgctttct cctgtgcctt tcgctgctgc tcctgggctg ctgggcggag ctgggcagcggg

SEQ ID NO:15

10

cgaaacgatt caaaacctct ttactgccgt tatttgctgg atttttattg ctgttttatt tggttctggc aggac

SEQ ID NO:16

aaatgagcag daactctcct gtgaatccaa aagactccgg agaagaggct gcaaaagtta gcgcggagag acaccatgga tactatccag ctgtacctgc ctgtgcaaga cacacgatga gcatccagga ccagcagggg agagtccctg ttgggtccgc tggtagcacc caagaagacc gagacaatac ggagtctggg ggaggcttag tgcagcctgg atgacatgtc gcagccatta atagtgatgg ccttgtatta ctcctgcaga gagacgattc atcatctcca ctgtgagctc aaactaaaac ttcccttccc gaggacacag acacagagtt gtgaaat tgaatacgaa ggagttggtc tctgaggtct

SEQ ID NO:17

20

ccctgggagg caagccagac tgcattgctc atcataataa aatcagaaat gcagcaaacc tttaaaaatat acatccattc ttctgtgagt ttcactcgaa gagcagcgtg tcactgcgga ggaactggcc caacagtgtg actacaccca agaagtgtgc ggtctcagag cagacctctg cccagggtac tgcatgactg cgccaaagtg agctcaccat ggctggacct cttcacggac acaggctgcc cctcttgctt tggcaaccag gtcacatctt aaagcatatt ccagacttgg

25

SEQ ID NO:18

30 tggaagcaag agggagtatg ctaacttcat

ρ

SEQ ID NO:19

atcaattaca agaggg

SEQ ID NO:20 atgaagttmg catactccct cttgcttcca ttggcaggag tcagtgcttc agtkatcaat tacaagaga

5 SEQ ID NO:21

aattcttaat taattattgt ggtgtcacaa taacttttc

SEQ ID NO:22 cccccggat ccatggccgc taaattcgtc gtggttctgg ccgcttgcgt cgccctgagc 10 cactcggcta tggtgcgccg caagaagaac ggctac

cactoggota tggtgcgccg caagaagaac ggotac

SEQ ID NO:23 cccccggat ccatgaaact cctggtcgtg ttcgccatgt gcgtgcccgc tgccagcgct

15 SEQ ID NO:24 cagtgtacgg cggctcgagg cagaagtccg gacgcata

caaaccggac cgtttacggt cccagagagt ttaagcaacg ggatcaggat agaagtctga agcttcacga ttccaagcat tgcgattttg gtcggtaccg tgtttgcacg aaccatatg tgtagcattt aagtgtcacc gtcttctcag gaattcatta tcagcacgga ccagatttct tegttagegt ttcagttcgt cactctcgat tcgtagagcc tcgttgcgtt tttcgaagca cagaacgcat cccttccgct ggtattccag agtgcaaccg SEQ ID NO:25 ccatggatcc 20

25 SEQ ID NO:26 ttcacaggca ggtttttgta gagaggggca tgtcatagtc ctcactgtgg

agatgatacc taatgttgac ttacaaaatc cttcttccca tagtattgca ttctgtgtac attggatttt aagtggccct ctctgtttct cttttagtga atttgttttc tttgcttatc catgttggca ttgccttgga aagcttctcg tgaaaaccaa cccaattagt aagtactgtt cgaagtgggg agctaatggg aatattaaaa atattttaaa atacctccat ttatgattgt tgcaaaagac atggctaaag tctgttaagt SEQ ID NO:27 ggatgggaaa tttatcattt ggaagagat 30

SEQ ID NO:28 atggctacag gctcccggac gtccctgctc ctggcttttg gcctgctctg cctgccctgg cttcaagagg gcagtgcc

5

SEQ ID NO:29

atgaggtct ttgctaatct tggtgctttg cttcctgccc ctggctgctc tgggg

SEQ ID NO:30

10

atgc acctgagaat ccacgcgaga cggaaccctc ctcgccggcc ggcctggacg cttgggatct ggtccctttt ctggggatgt atcgtcagct

SEQ ID NO:31

atg gccattagtg gagtccctgt gctaggattt ttcatcatag ctgtgctgat gagcgctcag gaatcatggg ctatcaaaga agaacatgtg atcatccagg ccgagttcta tctgaatcct gaccaatcag gcgagtttat gtttgac

SEQ ID NO:32

20 agggggatc cccggagacc ttcgggtagc aactgtcacc ttgatgctgg cgatcctgag ctcctcactg gctgagggc

SEQ ID NO:33

atg gtgtgtctga ggctccctgg aggctcctgc atggcagttc tgacagtgac 25 actgatggtg ctgagctccc cactggcttt ggct

SEQ ID NO:34

ggtgctagcc aaacagctgt tgttgagcga attgttagaa catcttctgg agaaggacat ggccaaagtg ggcagtttca gccagaatgt gatattggga gtgtgtggca tgcatcctca tcatcaggaa actctaaaaa agaaccgagt 1 gaacgattta aggagcgaat actactggta aactaatgga agaaatctgc tgcaccactg catcaccttg gaaatgaggg agctcatcca ggaactcctc aacttgctgc ctaagagggg 181 241 19

39

241 ggaactcctc aacttgctgc ctaagagggg tccccaagct tttgatgcct tctgtgaagc 301 actgagggag accaagcaag gccacctgga ggatatgttg ctcaccaccc tttctgggct 361 tcagcatgta ctcccaccgt tgagctgtga ctacgacttg agtctccctt ttccggtgtg

gatgtctgag aagcaaacat tcagacatga cctggcacag tttctgttct agagccagcc gtg:tcagtt tagagtgaga cgcagcctgt gagcatcaga gtettttee aggatcccct gggtgcgagg cgaggttcct cacgtggccg cacctctacc aagtggaagc tctctgcctt caagtgaagt ttcagtactg tctcaaagat tcaacctgca tatctctcc atgttattgt gggatgtgga ttctatgtga ctgcacaccg ctctttgaca acgccaactg gaggtgctat ctctatgaga ccatctatgg aattttatca aacactccct gaacaccaaa gtgcgttccc agtgccatac aagagcctga ctcaaggcac ttgttcctgc ttgtgatatg gtgcccatca ctttgcctgt ggctctgaag cccatggaag gcctcgtggg ccctgcttca gcccacgcgc tetgtgeege cccatattt cattaqttaa ctttttgtcc tttcattcaa atctgtctag aggttatgct catccacgcc ggatgcacgg ccacctctct gcctctcttg gcacagttac gtggaggcg cgtggcctag gacgtccatg gcctgccgtg gatactgtgg tgcactcctg cgctctggag aggatcggga gactctgtaa atgtagttct actgcagcac gctgcagttg gcttttatta actgaaatga ctattcccac tagcttatct ctgccaccaa ggttgaccaa caagatggaa agaaccacgc cggtaaagaa aagttgccga agatgagact ccgccaugcg ctgagcgggc acatgatgtc acctccccat ttgatcttca cttgtttcct ctgtgaagcc catataaatt atceteegee attatcttt gaagcggcct ctcgtttggg ttgatctttg cttttatttc ggtttttcag ttetgttcae tgetgecaee actggaattt cttcatccag cctgtcgaca tttgggctat gcagaattt ctcgcatggt ggtgaagcct gtctcggcct gcacttatca atgtctgaat tttcattcat gagetttgtt tgtttcccta aaaagcttcc caaagatgta tattatgtgg gatagageet agctcctttt gaatgtttca gggagagagg tttagtgatt taatgtagac gcagtgtcca acactcttgg ttgtacctta ctctccctta ctgttggctg aaagggactg caagtgtttt atctcccaga tggcactcct agctccaaga caaaaatgtt agaagctccg ataggttgca gagagaaaga tcttcaagct aagagaaact tctgccttca tgcctgcctc agatttctag tcatttcatc tgtaaatgct tctcctttga ggctcttgct taaggtgaac ggatgtggga ctattgccag tagatggcac tgtccaaatt cagtggaagg acaccetece caggaggtgt actccaggcc gtgtggttat ctatcttta gcaccttgct tgccttgcat ttgtttctct aagaatggtg gtgcaaggag acgtccatct caggaaatgc gggcacctcc tcctgcatcg aaactgctcc cccaagccta cagaacaaac cacttcactg ctagtcaccc cagctggcat ccctttaca gatggtcctg tatgcggcta agtattaaac ctgatcgtgg ttggttggac tacaggatca agagtgatgc ggtacatcga tgagtgtggg ccagcttttg agaccatctc gactagagac cagcgccacc cagccaacat cttcgttgg gaaggccagg ttcctgagaa ttctttacct acatgctggt aattccaccg tgttcccagg cactggacca qccccctcag tgtaaacaca gtttggaagg cttcatatct agtcacggac tgtggatggg tggatccctt ccacagtact ccagactgca tgagtcctgt agacaataaa aacacacttc gagcaatgtg 2461 1021 1081 1141 1201 1261 1321 1381 1441 1501 1561 1621 1681 1741 1801 1861 1921 1981 2041 2101 2161 2221 2281 2341 2401 841 481 541 601 661 721 781 901 196 2 25 30 35 S 10 15

ctgttcttta gcccaggctg ccggctaatt cgtgagccac tettettee gttgtttgct atgagccacc acgccattct cttgatctcc gaactcctgg cttagaaccc tatgagaact tccagcccga ggccttctca tcttgacaat tattgtattt catttctgtc catcatgtac agggaagaag acaatataca ggcattttta ggtcuctage cccaacttcc tegetetgte tecegggtte gccaccgtgc ccgggatggt ggattacagg gcttacaggc ttggttttta attttttttg gactggtctc caccccctg caccttggtt cttgctaagc tttacttgtt aaacttccat ccttttcctg cttcctcaag tgtcaccaca ttttattcac tgaataaag gttccttttg gaagtatttg tttatccact catctcccat tttgtttggg gtgccaaata tctgaaactt agacagagtc aagctccaac acaggtgccc accatgttag caaagtgttg aaatagctgg aaaaaaaa atgtttctca taaatctgat gtaaactctt atggtgattt tettteetge tggtgtcctt cttctgtgtc atatgctact accttgtttg tcaataaata aaatgtttgt atcctacctt ttgataaaat gtgttttgct acatagtatg acttccttca ttgccatcgg tccctgctta tttttttg agctcactgc agctgggact acagggtttc ctcagcctcc cagagtagtt agttctcact tcagcctctc aattttgtgc gagcagtcta ttttaaaaa tgtttaaata catttaaaat acctcattaa ttgactcagg gaggcagaa cagtttatat ggccttcatt gtcaaggtgc gtttttccat gtgtgagata ctttgggtcg ggattttgta aaaaa tttagtagag tctcacataa ttttaaggtg tettecegee gaaaccttta tttttttt gcacgatctc ccttccaagt atccacccac caagaccttt tattaccct ttccccttga aggatttggt ctcagcttct cccaccactc ttcagtccct gcatttattg ggctgtaaat gcacatagtg cttagctaca tttgtacttt gacatatcat gcctgtagta caagtcctgc aaaaaaaa ccactgcca tttcaaccta cctgcctcag tgacctcatg agacttttt gagggcagtg tttttgtatt tgcgcccggg gtttttgtt cctcaagcca acacctggcc ttctgctttc tccagtctgt ggttcttgcc gtggtctcct gctctccacc tcccgggccc tcccctacca gcaagccaac cggattatta tcccctcatt caagtgcttg gctccttgaa aaaatgctag gccttgctgc cctttccact tttattaaaa 2581 2641 2701 2761 2821 2941 2881 3061 3121 3001 3181 3241 3301 3361 3421 3481 3541 3601 3901 3721 3781 3841 3961 4021 3661 S 120 15 20 25

SEQ ID NO:35

gatgttggca tcaagctgtc aggagctcag ggaagaggtc gagcaccaga aagcccagca gagaaaggaa tccaagaagc caccagctaa actgcagcca ccacccagag gacagaggac aagcagaaga ggcagaggag ctactcctca gcatgctctg gatggacatg gccatggcag gttccagctt cttgagccca 1 tccagatcat ctgtcctcac caccaaggcc atggtgtctt cagcgactat ctgcagtttg taccagcagc atggccgggc cctgggaaag tttcttcagg atatcctctg tecetgtact cagotaacaa gtaaccactg acaggactgg gagctggaaa tcaggttcaa tgctcccttc cgagctctgg aaggctggct aaagaggcgc 121 181 241 361 301 19 30 35

gcaagaacte acatecaget tetgeeteet etgeaactee cageactete etgetgaett 421

481 acaaataaat gttcaagctg t

SEQ ID NO:36

atgaagctgc ttgcaatggt tgcactgctg gtcaccatct gtagcctaga aggagctttg

atgaagctgc ttgcaatggt tgcactgctg gtcaccatct gtagcctaga aggagctttg SEQ ID NO:37 gttcggagac 10

gttcggagac

atggeett gecaaegget egaeeeetgt tggggteetg tgggaeeeee geeeteggea geeteetgtt cetgetette ageeteggat gggtgeagee e SEQ ID NO:38 15

atgeegegee tgtteteeta eeteetaggt gtetggetge teetgageea aetteeeaga gaaatcccag gc SEQ ID NO:39

SEQ ID NO:40 atgacagc atcacttgtc gttttaccat cgctttggtt aatattaatt atttttactg caccctatac tcactgt

25

ttgtcagagg gactgtctca catgccagga gaagctccac ccagccctgg acagcttcga tctggactcc cagagcatgt gaatgccccg aggctggtga atgaaagte etgetttgtg acetgetget geteagtete ttetecagtg tgtteageag agtecggage ttgttecagg ageaggaaga geeegageet ggeatggagg agtgcgaaga gaaggtcttc cccagccccc ggagctcttg gcagctcagc cctgccgccc ctctaccagc cgagagcttc ggagatgcag catctgcggc gatggagcag aagcagctgc agaagagat tgcatcctcg atgcaccaag gtcatggcca SEQ ID NO:41 ggcggctgct cctggaggtg 30

aattcatgaa gtgggttact ttcatctctt tgttgttctt gttctcttct gcttactcta gaggtgtttt cagacg SEQ ID NO:42

5

atgaagtggg ttactttcat ctctttgttg ttcttgttct cttctgctta ctctagaggt gttttcagac gc SEQ ID NO:43

SEQ ID NO:44

10

gctggtgtta tttagaatac acaaaaacca gagaaagaaa acacactctg tecetetgte atacacacae acacacae acacacaca acacacae cccaggtacc gcagcctagt agaagaataa ctcatcgcag 61 tgttgccggc cagtcaccca gttgagggat ttgaatgaca tcataaccct caagagggta tggcaaaggc aagtgtacat tataaatagc aaaacagctg gcttggacca ttgtttgccc tagtetttge aacaggaaaa acagtgacag cagctcggca gttatatgga aaattcaaac tggggaggg ttgggtttta tctatttcag agtggtgggg ttgctagcca gcacacagac agaggtttga cttctcccag agaaaaatgt 1 gaattotoaa 121 181 241 301 361 2

tcacccgagt aaaattgtag agtttgctgg ccctccata ggacaagggg atgctgcctg gaccaatcca ttggcaactg cttcagggtt gccctaaagg cacccctcac accccagtca atgagacagt aatgagatca aatctacttt gatgtatta gagaaaagt cgttctagga 421 481 541 20

gctctagaat ttctatggca ttagcagacg aagggaaaaa gtaaggtgta ttgttttaaa taatgcattt gatccttatt acattcttt tttaatactg tcagaaaca tcagtaacac aaggtatgat cttctttt gcgaaccctg ttaaaatgca cccatattcc agaaagttaa agaggtctt atcgtccttt teteccatec catgcccagc tttaatgtct tgctctagtt ctaaaattgc tttgtgtctt aatgcttctg gcataagtat 721 601 661 781 841 25

aatccctctt ctgtttgcta aatctcactg tcactgctaa attcagagca gatagagcct gttatttgtc accatgccca attttgtgta caaatgttcc tcctgtctac agtgtctgtg ttttgtagat aaatgtgagg ctttgccaga agagggagag agagagaagg tactcaataa cccagctgtt aaaagtcct 1021 901 961

gctctcaaca tcattttcag caaaattgaa atgtgacatt ccaacccaat gctaaccaat ctttttgctt cattattcct cagcagtett gcgcaatgga ataaagtcct tgggaaaaat ctctggattt tcagaagcaa 1081 1141 1201 30

teteccatet actttgtact ctgcttttgt

35

1381 1441

tgaaatagga ccaggggaat

cctccaaaga tccatccaca

agggaggact

aaaggggttc

tgttggatgc cgagacacgc

tggggaattc

ttgaagtcat

ttcttactct

aggtaaatat catctttcat acccataaag aaacacaggt

gatttcttga

1261 1321

gtaataatag

ttccgtagta aacgtttcta

PCT/US2003/041269 WO 2004/067743

atgaa gtgggtaacc tttatttccc ttcttttct ctttagctcg gcttattcca ggggtgtgtt tcaaggagga ctgattaatg acagtcgtgg ctcctcgcat ccctctggcc ggggcctctg caaacaggac atgactcgct ccgccttggc cgagctgctc ctcgccccc gggagagaaa ggccggctgc aagaacttct tctggaaaac cttcacctcc gectecatea getgeteett egeeggaeag agagaeteea aacteegeet getgetgeae 1 atgaagatgg teteeteete gegeeteege tgeeteeteg tgeteetget gteeetgaee gggcactggc tgcctgcggg gaagatgaaa cggcacaaat cccggctgaa gctgtcatcg ttgctgtttt gccattttcc aacagcacaa attgtttata aatactacta ttgccagcat tgctgctaaa gaagaagggg aggotgotoc agccagcact ggtoctatgg actgogocot ggaggaaaga gagatgooga atgagatt toottcaatt tttactgcag ttttattcgc agcatcctcc gcattagctg atgaag ccaattcaaa aactcctagc tggccttatt ctactgactt cgtgcgtgga acaccatgtc cttgcaagga ggagaacgag gctctggagg aggagaactt ccgccagcgg ctgggtagtc cccactccc gtgaagatgc ttcaccagct ccacgccgac ctagagcggg gctctgtgcg ggcttctgtg tggcgctgtg cctgctcacc ttctgtccca tcttgggcag cgcgctgctg gatttcgatg tatctttgga taaaagagag gctgaagctt gtactctatt gtcaatgcac tecetetgtt ctqtcggacc tcctgcaggg ccgaggacgc eggtaceege tgeagggete ctccaqtcaa cactacaaca tttagaaggg ctcttctacc aacctgaacc gggttctcat aatcgtggga gaaggaggac SEQ ID NO:49 SEQ ID NO:45 SEQ ID NO:46 SEQ ID NO:48 SEQ ID NO:47 gttacttaga ataacgggtt tegtegaga aaattt 1561 1681 1621 181 241 121 301 361 61 10 15 20 25 30 5

gaggcggcgc cgagctgcgc

cggcgggcat accgctaccc

ttcgagtacc გეებენგნგნ

ccaggagccc cggcatctcc

gcgccgaagc agcaagagga

cggcggctgc

tggctcctgg

atggccgggc gagggggcag

ggctgttgga tggcgtccat gggtcgaage gcaggattta cacggtgggg ggggaacgag gaacccagat agtgtacgtg aattgtggat ggatattcct atatgatgag tttccaaagc tgcttggtac ctgggaggat gccaccatgt tgagatcacc aggatttgtc aataqaccac teettacage aaactataaa atgggaatga ccgacaaccc aataccagaa tgccttccct actggtttgt atatgaagaa tggataggat attggattat ccaattatcc atgacgccat gaggagttaa accccaatcg gcaactgttt tgaagaccta caacggtgg ccgtggaagg aaaggaaaga aagtggcagt taaatttt gctgcagtgc accgccatca gctgtcattc atcgagctgt gggaatatgc ctatgcaacg attcacatca gaactcaagg tttccagacc ctgttgaaaa gaccttgtgg tcctccccag gccatgtctg gatggaacca taccttagca gaagagactc cagatacacc ataacaaaga tcattttctg gccaccatct tcagaaactt gctcctggtc ggcccagtac cagtacccgc gaaccggaac taaatacatt aaataatcat tgagaccaag tcagcctggt ccatggagga cgaatacagc tttcaacccg ctaccttgag tgaactggag cagctttgta agacttcaat gttcccacct aattgcgaat tggtgattac ctatctggca gaaaatgatg tgtccgtgtg aggcagcgtc agggccggga agcctgaatt acctgatcca aaggtggtcc tcattttctt gaatagatct agcttgctcc ctgccaatct gtagtgctca catactcttc atgatgacag gagggatgca ccctcattag ccgcaaagga cagetecagg gctgtgagaa aaggtaaccc gggttgattt tggaatggtg gaggcgctcg ggctttgaga cgcagcttcg gagcctggtg cgagaactgc acaattgtca aatgcccagg aatgagaaag acgcggagtg ttgtgcttt ttggcccggg cgcaagaatg agcgtacctg cgagaccttc caaaacacaa gtggagetta aacaaaact cttacagcct gatgttacat gaagaattga cctgctgctg 5 10 15 20

SEQ ID NO:50

atggctctct cactcttcac tgttggacaa ttaattttct tattttggac actcagaatc

SEQ ID NO:51

atgaacaaa ctagcaatto togotatoat ogotatggta ottttoagog caaacgoott cagaotocaa agcagattga gatcaaatat ggaagottot gocagag

SEQ ID NO:52

30

a tggtcagtgt gtgcaggctc ttgctggttg ctgccttgct gctgtgtttg caagcacagc tgictitctc tcagcactgg tctcatggct ggtaccctgg aggaaagaga gaaatcgact gaatgcagct gattgacacc cttgcaagga tgaagcggga tctggggaga ttaaactgtg atctcaggcc actgaggacc aacatcctaa agagcatcct cctacagctc accagagata 35

## aattccaaaa gaggaaatga

SEQ ID NO:53

121 ttgaaatgtg actttgctct aacatctccc atctctctgg atttcttttt gcctcattat 241 tottccaact caattattta agatctgcct ctgtgacttc ttgaagataa agatacacat 1 tgtgttttgt agataaatgt gaggattttc tctaaatccc tcttctgctt gctaaatctc gcgatgggga 61 actgtcgctg ctaaattcag agcagataga gcctgcgcaa tcgaaataaa tootgoocad caattoattt coagactttg tacttoagaa catg 181 301 S

10

atgte aggeeegagg aegtgettet gtetaeegte ggetettgta etagtaetge tgagteteag cacttoggca ctaggg SEQ ID NO:54

SEQ ID NO:55

agefvhmarn afsaggpgkg sidrgvshfv vqlladttlr epgrarnmts dietayvggs isavtrddlv yvdagpiyet rleplelvnn trapssvths tlqvlmgggg grglgnsifc svlhhdlprt sfnphlcnvl qpvgsagsfg snshfaavgv fthpdmfpvs igrqvlaqnq afqhayssts lfgisascvp atresilyqg gemmhavafg lvskafvpss aafnasgeal lesqvitved tgvahaelvd hvqvafpvpp psqirvatap aglggnlfct dlhsrpdafi tkltdiketl sslafksthg atesqvlktm mspaaqlaka aarstcmtrl hqlvipkppp thpnyeqtlt nryrwmesca lsdeevarak nglkssllmn ptmvavgedl rratiafeae lhpsrmvag palteeeiae rvaealvakp dtireyfaty mysrlytnvl 15 20

(prints) SEQ ID NO:56 25

gccattgcat acgttgtatc catatcataa tggagttccg atcatatdcc tegetattae actcacgggg cccgcccatt 61 cogotacact tyccayogco ctagogcocg ctccttcogc tttcttccct tocttctcg 1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga tgttgacatt ttacggggtc attagttcat agcccatata cccaacgacc gggactttcc catcaagtgt cggttttggc agtacatcaa tgggcgtgga tagcggtttg gcctggcatt gtattagtca tatattggct catgtccaac attaccgcca atggcccgcc tggctgaccg cttggcagta tcaatgacgg taaatggccc ttcccatagt aacgccaata gtacatctac ccacgttcgc cggcatcaga ttggctattg aaactgccca ctacttggca cttacggtaa tagtaatcaa tgggactttc atgacgtatg tatttacggt cctattgacg cgttacataa gacgtcaata tatgtacatt tagttattaa catggtgatg atgggtggag aagtacgccc catgacctta 301 121 241 181 361 421 481 541 501 30 35

atacacccc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt gatetecaeg cttgctccta aatccataac cttcaqaqac aattcacata agcttccaca cacggctgac tgttgtattc cagtgtagtc gccgcacaag gactaacaga cctggagacg actcgatatt caacagctta acttggccgt attgttaggt gcaaaaacga ttatgagaaa ctttaaagtg aagaggaaaa tccgcggccg ccgcctatag cttggggcct ctttatctgt agccgacctt acatgatatg tccaatctca ctcgacacgg gccatgggtt tatctattcg aaaatcaacq aatcgccctg ttgggacaag cttaggcatg cggactaggc ggctgcaacc aggggatcgc gtaggcgtg tgacgatact ttccattact atactctgtc attatttaca catagogtgg teggeagete ncaccagtgt attgggctcg ttttggcacc caaatgggcg cacacccctt tggctcttat gcatgctata ctgtttttgg gtagcggcgg gcagctgagt cggtggaggg atagctgaca ccatgtgcga taaaacgact ctcttaccga gaatcgaccg ggcatgctag atattogtga tcagcaactt cgatccagcc gacgtaagta gtgatgctgg ctaaaagcaa cgtcagatcg ctgttactct accaatttct taagtcgagt cggcaaagga aaaatcagcg aacttgttaa qtcctgccta tcatgctgct caacagttcg agaaacaagg ggaaattatg acttactcgt ggtcccattt ctatatgcca ttttattaaa ctcttctccg tcatggtcgc gagcgtggag gaagatgcag gtgctgttaa caccgtcgga atgcccacca accagacata acaccgggac gaattacact aacaaaacat aacatcaaac gtataccgtt gactggtctg aaagctcatg tgggagtttg cccattgacg tttagtgaac tgccaagagt aaaactctca cacggtcgtt ctcattgtca tggtactggt tggaaaccta aagaggctga aaaggccgaa tactcagcgt acacccaaac gttcatqctc gacttgaaaa cgttttgata aacgtactct tacgetttgg acaagggaag ccctattgg tototattgg cggacatggg ctccagcggc tacaggatgg tgcccgcagt gcacagcaca gtctgaaaat agcggcagaa tecegttgeg ათნანანანა tttctgcagt attetgeece gagactagat atgttcaaag ttgacgtcaa acaactccgc gcagageteg tccatagaag acttgactgt tgtacttgtt caccacaccg agtegeacta agcggaaac atctcgctct acctaaaatc tgaaattcga ggattccccg gaagctgggt tttaggctat aaccttccga gcttgcgggc cagaaatcga cageteagag ctacacaata cacacatggt gccacaacta attgaccact totgtattt gtacgtgttc cagaggtaac ccgtcccccg ggtcccatgc tagggtatgt gacttaaggc tegttgetge ggaacggtgc attggaacgc ccagacttag ccatgggtct ctctttacca gccacqcatt ccaaagcgag tccacaaaga tttcagagca ctccacccca aaatgtcgta tgttttgacc tgcgagette taccgagtaa aatccgttga gtctatataa cgatgcccat cagacctagg actcaaagac tattgtataa accacccgtc agattgaaga acttacctgt taacatgttg gccgaacgag ctaacacagt ggcattctgg aaaatttatt attgaccatt atggctcttt tgacacggac cgaatctcgg tgataagagt tacaacaacg acagtggagg gcagatggaa tgagcagtac ttacacgact atttccaagt ggactttcca acggtgggag actctatagg tccgagccct gccgtggcgg ctgttccttt acgttggctt aacctgccaa aatcgtcacc tcgggcaata cttatggtat gegtteeege gcgagcattc ccatggtata gtacaatatg tcatctagtc ccatccacgo tgccaaattc actcattgtc ctagcaacta aagcgaatgc cacttccagg ctacqccata atgetteaac gaagttttgc ctactagete 1021 1081 1141 1201 1261 1561 841 1321 1381 1441 1501 1681 1741 1981 1621 1801 861 1921 2041 2101 721 781 961 901 2161 2221 2341 2461 2521 2701 2281 2401 2581 2641 2761 2881 40 10 15 20 25 30 35

gttaaaattc aggccaattt attgctattt tggaattgtg tgacctcgag gccttgcagc gcccttccca cggcaaaatc ttggaacaag ctatcagggc acattccqtc atgcatatcg tttgaagcta atatttcgcg tcccctgagc acgctaaacc cagcaaacad gtaaagcacc aagcgatccc taaacgctaa tttcgcccat accaagaccc gcaccacacc gttaaataaq acacaacata actcacatta gctgcattaa ttagctcact ctctagaact ttaatgattt ttatqattqa tccaagaagt attactggta ttcctgacgg ggtaaaaatc cactggccgt aaagtatatt gcgcgcaatt tcctgtttac atacagtaaa aatgcaggta agtgtaaaag gagtgagcta ttgttccagt gaaaaaccgt ccaccttaac acgagaatta atggctggtt ataggtttta gagggtcatt ttataaataa ccttcagata actgccgttt ttctgtaata aaatggcatc tgtcgtgcca gtatgttgtg gtggcggccg tegatacege cgcaccgatc ttaatatttt aggccgaaat tcacaattcc aaaggtgcct cagcaattaa ccttgaatac taatgatttt gtaatgtgag attacgccaa tacgcgcgct aataatccct taattgcgcg acttaaagct aacttaatc tctgccttag cttatcccta aaggeteace tgttatccgc ggtgcctaat tegggaaace tagtgaatga atgaatcccc tatgatcccg tgagtcgtat tggcgttacc agatgtgtat ctcagggtca aacttactca ccqataaaa aataaaatag gggttatcaa agcacttgtc caggataata cactccctgt aactaaccaa tccgagcagt tggaatactg tcatcacqat aataatcaac tagtgagggt gttggtgacc aagagaacca cttccggctc tatgaccatg atcaagctta cgaagaggcc gggttgagtg gtcaaagggc ctgttataag aaatggttgg taataaaac aaaaattag ctccaccgcg attgtaagcg tttaaccaat cageteattt tgtgcttaaa ggatagacat aaacagggaa ctgcaataaa cacaaaggct gcgctgaaat ttgttccctt tgtgtgaaat ggaattcgat cgccctatag gggaaaaccc gcgaatggaa gaccgagata tcatatgaca gaaaaaccta aaagagcttg atgeeetett gaaataacta ttgcattggt atgctattca taaagcctgg cgctttccag cgggaaaagc ctgatttaac ttattgataa atgaacttga aaattgatcc ctgagcactg catcttgtca aggaaacagc aaagctggag ggcgtaatag ggactccaac ttcatcagtg ttgaaagata tcatcgatag tactcagtgc acactttatg cccaaattgg agcggataac aatttcacac tttattgagc atcgtaaaaa gccaaccatc ggctatcaat cccaggettt caaaagaata atgatgatga qaataacatc atttggtgac ataacaccgg gcactactat ttetteetge cgatcagctt gcataaagtg geteactgee ggtggataca aaagggaaca ccgggctgca gtacccaatt cgtcgtgact ttcgccagct cattagggga aataatccaa atccagccat ctctttgttt aataaaatta agctgtttcc tccgggatct cttgattata attggttgga cattgtcagc tttgttaaat tactccggga aacatgaagg aaaaaacaat acaacaaaa ctcgttgacc agcctgaatg taaagaacgt agtggatccc ttgaggggtt gatgcctatc taatgggcca ttgctgatga accaccagcc aaaggcaaat ttagtggcta gtaatgaaaa gtgctggatt tgatgtatac tcatggtcat cgagccggaa attgcgttgc ctttaaaaat tgaacattat aatgttaatt attaagttaa cattaggcac 60006666666 aggaaagctt accgcggctt aatcttcttt tgcctacatc aggtaaacgc aaccctcact cgttttacaa gcgttaaatt agtccactat gatggcccac ttaccaaaat caatacatgo acatccccct acagttgcgc ccttataaat 3301 3421 4081 4741 4801 4981 5041 5101 5161 5281 3181 3241 3361 3481 3541 3601 3661 3721 3781 3841 3901 3961 4141 4261 4501 4561 4621 4681 4861 4921 3121 4021 4201 4321 4381 4441 5 2 25 30 35 10 15

tcactcaaaq ggcgctttct caggattagc ctacggctac ccataggctc aaacccgaca tectgtteeg gctgggctgt tegtettgag cggaaaaaga ttttgtttgc gagattatca cttttctacq aatctaaagt acctatctca gataactacg cccacgetca cagaagtggt tagagtaagt cgtggtgtca gcgagttaca cgttgtcaga ttctcttact gtcattctga taataccgcg gcgaaaactc aaggcaaaat atttgaatgt attactttt ggcgctcttc cggtatcagc tggcgtttt agaggtggcg tegtgegete ccactggtaa ggtggcctaa cgggaagcgt ttcgctccaa ccggtaacta gcggtggttt gaaagaacat cagttacctt ntcctttgat ttttggtcat gttttaaatc tcagtgaggc ccgtcgtgta taccgcgaga gggccgagcg aacgatcaag actcaaccaa caatacggga ccactcgtgc gccgggaagc ctacaggcat cactgcataa gttcttcggg gtcctccgat caaaaacagg gcggatacat cccgaaaagt tactcatact ttgcgtattg gataacgcag gaagaattga cgctcaagtc ggaageteee tttctccctt gtgtaggtcg ctggcagcag ctgctgaagc ctgcggcgag tgcgccttat ttcttgaagt accgctggta tctcaagaag cgttaaggga gcctgactcc ccagccggaa teeggtteee taaaaatgaa caatgcttaa gctgcaatga attaattgtt gttgccattg gttatggcag actggtgagt tgcccggcgt attggaaaac togatgtaac tctgggtgag tgtctcatga cgcacatttc agctccttcg aaatgttgaa ggtcgttcgg cttatcgcca gagaggcggt agaatcaggg ccgtaaaag caaaaatcga gtttcccct cctgtccgcc tctcagttcg gcccgaccgc tatctgcgct caaacaacc cgaaaactca tgctacagag aaaaaagga ccttttaaat tgacagttac aataaaccag aaaagcggtt atcactcatg atccatagtt tggccccagt gcgcaacgtt ttcattcago ctttctgtg catccagtct gagttgctct agtgctcatc gagatccagt caccagcgtt aggggttccg ggcgacacgg tcagggttat ggttatccac cgctgcgctc gataccaggc gctgtaggta aggccaggaa acgagcatca taagacacga atgtaggcgg cttgatccgg ttacgcgcag ctcagtggaa tatttcgttc ttaccggata ccccgttca cagtatttgg tcacctagat aaacttggtc gcttaccatc atttatcagc ttggtatggc tgttgtgcaa tatccgcctc ttaatagttt tgcggcgacc ccgcagtgtt ccgtaagatg gaactttaaa cttttacttt taccgctgtt agggaataag gaagcattta ataaacaaat ctcactgact gcggtaatac ogccccctg ggccagcaaa accetgeege tccaacccgg ggactataaa catageteae gtgcacgaac actagaagga gcgatctgtc agagcgaggt gttggtagct aagcagcaga gggtctgacg aaaaggatct atatatgagt atacgggagg ccggctccag cctgcaactt agttcgccag cgctcgtcgt tgatccccca gaatagtgta gccgcaaaaa gtcatgccat ccacatagca tcttcagcat agtaagttgg tcaaggatct atttagaaaa 5401 5461 5701 5521 5581 5641 5761 5821 6001 5881 5941 6061 6121 6181 6241 6301 6361 6421 6481 6541 6601 1999 6781 7081 6961 6721 6841 6901 7021 S 10 2 2 25 30

ID NO:57 (pInMod)

35

CTGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG 50

	ないしかい できしかし	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCG	CICCIIIICGC	100
		TCCTTTCTCG	CCACGTTCGC	CGGCATCAGA	TTGGCTATTG	150
	TODDITATION	ACGTTGTATC	CATATCATAA	TATGTACATT	TATATTGGCT	200
	ことをしていたから	ATTACCGCCA	TGTTGACATT	GATTATTGAC	TAGTTATTAA	250
v	TAGTAATCAA	TTACGGGGTC	ATTAGTTCAT	AGCCCATATA	TGGAGITCCG	300
)	CGTTACATAA	CTTACGGTAA	ATGGCCCGCC	TGGCTGACCG	CCCAACGACC	350
	正正をここことここ	GACGTCAATA	ATGACGTATG	TICCCATAGI	AACGCCAATA	400
	GGGACTTTCC	ATTGACGTCA	ATGGGTGGAG	TATTTACGGT	AAACTGCCCA	450
	CTTGGCAGTA	CATCAAGTGT	ATCATATGCC	AAGTACGCCC	CCTATTGACG	200
10	TCAATGACGG	TAAATGGCCC	GCCTGGCATT	ATGCCCAGTA	CATGACCTTA	550
2	TGGGACTTTC	CTACTTGGCA	GTACATCTAC	GTATTAGTCA	TCGCTATTAC	009
	CATGGTGATG	CGGTTTTGGC	AGTACATCAA	TGGGCGTGGA	TAGCGGTTTG	650
	PCTCACGGG	ATTTCCAAGT	CTCCACCCCA	TIGACGICAA	<b>JGGGAGTTTG</b>	700
	TTTTGGCACC	AAAATCAACG	GGACTTTCCA	AAATGTCGTA	ACAACTCCGC	750
15	CCCATTGACG	CAAATGGGCG	GTAGGCGTGT	ACGGTGGGAG	GTCTATATAA	800
}	BUTUE ADADADE	TTTAGTGAAC	CGTCAGATCG	CCTGGAGACG	CCATCCACGC	850
	TGTTTTGACC	TCCATAGAAG	ACACCGGGAC	CGATCCAGCC	TCCGCGGCCG	006
	GGAACGGTGC	ATTGGAACGC	GGATTCCCCG	TGCCAAGAGT	GACGIAAGIA	950
	CCGCCTATAG	ACTCTATAGG	CACACCCCTT	TGGCTCTTAT	GCATGCTATA	1000
20	CTGTTTTTGG	CITGGGGCCT	ATACACCCC	GCTTCCTTAT	GCTATAGGTG	1050
1		TTAGCCTATA	GGTGTGGGTT	ATTGACCATT	ATTGACCACT	1100
		TGACGATACT	TTCCATTACT	AATCCATAAC	ATGGCTCTTT	1150
	CCCCTACTACTA	TCTCTATTGG	CTATATGCCA	ATACTCTGTC	CTTCAGAGAC	1200
	TGACACGGAC	TCTGTATTT	TACAGGATGG	GGTCCCATTT	ATTATTACA	1250
25	AATTCACATA	TACAACAACG	CCGICCCCCG	TGCCCGCAGT	TTTTATTAAA	1300
ì	CATAGCGTGG	GATCTCCACG	CGAATCTCGG	GTACGTGTTC	CGGACATGGG	1350
	CTCTTCTCG	GTAGCGGCGG	AGCTTCCACA	TCCGAGCCCT	GGICCCAIGC	Ô
	しじじしむなししよう	TCATGGTCGC	TCGGCAGCTC	CTTGCTCCTA	ACAGTGGAGG	1450
	DESCRIPTING TO THE PROPERTY OF	GCACAGCACA	ATGCCCACCA	CCACCAGIGI	GCCGCACAAG	1500
30	じじしじじしじしじ	TAGGGTATGT	GTCTGAAAAT	GAGCGTGGAG	ATTGGGCTCG	1550
2	こるでにしてでしない	GCAGATGGAA	GACTTAAGGC	AGCGGCAGAA	GAAGATGCAG	1600
	Gradinasar	TGTTGTATTC	TGATAAGAGT	CAGAGGTAAC	TCCCGTTGCG	1650
	GTGCTGTTAA	CGGTGGAGGG	CAGTGTAGTC	TGAGCAGTAC	TCGTTGCTGC	0
	CGCGCGCGCGCC	ACCAGACATA	ATAGCTGACA	GACTAACAGA	CIGITCCITI	1750
35	CCATGGGTCT	TTTCTGCAGT	CACCGTCGGA	CCATGTGTGA	ACTIGATATI	1800

TTACATGATT

2250 2300 2350 2400 2100 2150 2200 2450 2500 2550 2600 2650 2700 2750 2800 2850 2900 2950 3000 3050 3100 3150 3200 3250 3300 3350 3400 3450 AAAACTCTCA CAAAAACGAC AAGCTCATGA TAAAACGACT AACAAAACAT TCCACAAAGA TCGGGAATAC TGTTACTCTT ACCACACCGC ATCCGTTGAG TACAATATGC CATGATATGT TAAAAGCAAT AAGGCCGAAA CCTABABICI CTTACCTGTT AGCGAATGCA GGACTAGGCC TTCATGCTCA CATGCTGCTA TACTAGCTCA GCATTCTGGC TAATGATCCA TGTGTGTTGG AGAAATCGAA TGTTGTTTGC CCACTGTCCT AGGTGTCATT GGATTGGGAA GTACCTCTCT CICICICICI TGCTGAAGAA TIGATGCCIA TTAGAAAGTA TGTTCAAAGA GAATTACACT ACCGAGTAAC ACTTGACTGT CCAAAGCGAG AATCGTCACC CTTTATCTGT TATTCGTGAG ACGGTCGTTC CATGGTATAA AGAGGAAAAG CAGCAACTTA AGAGGCTGAC TCTCGCTCTA CCACCCGTCA TAGCAACTAA ATCTATTCGA TCCTGCCTAC GTTTTGATAT CTTGCGGGCG TAACACAGTC AAGTTTTGCG GCTGCAACCC TCTAGAGATC GCCAGCCATC GGTGCCACTC GAAATTATGA TTGTCTGAGT GCAAGGGGGA GGCTCTATGG CTCTCGGTAC CGGTACCAGG TTAATTATGA TTGGTCTGCC ATCACTTTAA ATTCTGCCCC GCCACGCATT AACCTGCCAA ATTGTTAGGT GGCATGCTAG GTCGCACTAC ACTGGTCTGA AAGTCGAGTA TTCAGAGCAA CGAGCATICI TTTAAAGTGC GGAAACCTAT TTAGGCTATA ATTGTATAAA ACTIGITAAT CTCATTGTCA CCATGGGTTC ACTTGAAAAG AGCTCAGAGC AACATGTTGG ACTTCCAGGC TTAGGCATGG CTTACTCGTG ACGCTTTGGG AGATCAGAGC CCTTCTAGTT GACCCTGGAA TTGCATCGCA GGCCAGCACA GGATGCGGTG CICICICICI TAACAAATGG CTCTCTCTCT GCCTTTTATC TAAGCAGCAA GTACTTGTTG CTCTTTACCA ACGLIGGCIL GAATCGACCG GCGAGCTTCA GCCGACCTTG TGATGCTGGC ACTIGGCCGT GTATACCGTT CGTTCCCGCT GGTACTGGTT GCGGAAAACT CTCAAAGACT GCCAAATTCT TCGACACGGA GGCAAAGGAG CACCCAAACA ACCTTCCGAG CCGAACGAGC TGCTTCAACT AACAGTTCGC TGGGACAAGC CAAGGGAAGA ACACATGGTT GGCTAATAAA ATCTGCTGTG TGCCTTCCTT AATGAGGAAA GGGTGGGGTG GGCATGCTGG CICICICICI CTCTCTCTCT GTGCCTGTTA AAAACTGATT GACCAAAGGT CAACAGCTIA CTCTTACCGA AACATCAAAC GCGACTCGCT GATGCCCATT TTATGGTATT TATGAGAAAG TCATTGTCAG CCAATCTCAT AAATCAGCGC SAAATTCGAA CCAATTTCTA AAGCTGGGTT AGACCTAGGA **SATTGAAGAA** CATCTAGICA ACTCAGCGTC TACGCCATAG GAAACAAGGT ACGTACTCTC ATCGCCCTGA TACACAATAA AAATTTATTC GATCACTICI TTTTTG CCTCCCCCG TTCCTAATAA CTATTCTGGG GACAATAGCA CTCTCTCT TUTCICICI PATTACTCA CATCACAACA **LTGACCCGGT** S 10 15 20 25 30 35

	TATTTGAACA	TTATCTTGAT	TATATTE	ATAATAATAA		8
	CCTATCCAAG	AAGTGATGCC	TATCATTGGT	TGGAATGAAC		3650
	TTAGCCTTGA	ATACATTACT	GGTAAGGTAA	ACGCCATIGI		3700
	ATCCAAGAGA	ACCAACTTAA	AGCTTTCCTG	ACGGAATGIT	AATTCTCGTT	3750
<b>5</b>	GACCCTGAGC	ACTGATGAAT	CCCCTAATGA	TTTTGGTAAA		3800
)	TTAAGGTGGA	TACACATCTT	GTCATATGAT	CCCGGTAATG	TGAGTTAGCT	85
	CACTCATTAG	GCACCCCAGG	CTTTACACTT	TATGCTTCCG	GCTCGTATGT	3900
	TGTGTGGAAT	TGTGAGCGGA	TAACAATTTC	ACACAGGAAA	CAGCTATGAC	3950
	CATGATTACG	CCAAGCGCGC	AATTAACCCT	CACTAAAGGG	AACAAAAGCT	4000
10	GGAGCTCCAC	CGCGGTGGCG	GCCGCTCTAG	AACTAGTGGA	TCCCCCGGGC	4050
2	TGCAGGAATT	CGATATCAAG	CTTATCGATA	CCGCTGACCT		4100
	CCCGGTACCC	AATTCGCCCT	ATAGTGAGTC	GTATTACGCG	CGCTCACTGG	4150
	CCGTCGTTTT	ACAACGTCGT	GACTGGGAAA	ACCCTGGCGT	FACCCAACTT	4200
	AATCGCCTTG	CAGCACATCC	CCCTTTCGCC	AGCTGGCGTA	ATAGCGAAGA	4250
15	GGCCCGCACC	GATCGCCCTT	CCCAACAGIT	GCGCAGCCTG	AATGGCGAAT	
)	GGAAATTGTA	AGCGTTAATA	TTTTGTTAAA	ATTCGCGTTA	AATTTTTGTT	4350
	AAATCAGCTC	ATTTTTAAC	CAATAGGCCG	AAATCGGCAA	AATCCCTTAT	4400
	AAATCAAAAG	AATAGACCGA	GATAGGGTTG	AGTGTTGTTC	CAGTTTGGAA	4450
	CAAGAGTCCA	CTATTAAAGA	ACGIGGACIC	CAACGICAAA	GGGCGAAAAA	4500
20	CCGTCTATCA	GGGCGATGGC	CCACTACTCC	GGGATCATAT	GACAAGATGT	4550
1	GTATCCACCT	TAACTTAATG	ATTTTACCA	AAATCATTAG	GGGATTCATC	4600
	AGTGCTCAGG	GTCAACGAGA	ATTAACATTC	CGTCAGGAAA	GCTTATGATG	4650
	ATGATGTGCT	TAAAAACTTA	CTCAATGGCT	GGTTATGCAT	ATCGCAATAC	4700
	ATGCGAAAAA	CCTAAAAGAG	CTTGCCGATA	AAAAAGGCCA	ATTTATTGCT	4750
25	ATTTACCGCG	GCTTTTTATT	GAGCTTGAAA	GATAAATAAA	ATAGATAGGT	4800
ì	TTTATTIGAA	GCTAAATCTT	CITIAICGIA	AAAAATGCCC	TCTTGGGTTA	4850
	TCAAGAGGGT	CATTATATT	CGCGGAATAA	CATCATTTGG	TGACGAAATA	4900
	ACTAAGCACT	TGTCTCCTGT	TTACTCCCT	GAGCTTGAGG	GGTTAACATG	4950
	AAGGTCATCG	ATAGCAGGAT	AATAATACAG	TAAAACGCTA	AACCAATAAT	2000
30	CCAAATCCAG	CCATCCCAAA	TTGGTAGTGA	ATGATTATAA	ATAACAGCAA	5050
)	ACAGTAATGG	GCCAATAACA	CCGGTTGCAT	TGGTAAGGCT	CACCAATAAT	5100
	CCCTGTAAAG	CACCTTGCTG	ATGACTCTTT	GTTTGGATAG	ACATCACTCC	5150
	CTGTAATGCA	GGTAAAGCGA	TCCCACCACC	AGCCAATAAA	ATTAAAACAG	5200
	GGAAACTAA	CCAACCTTCA	GATATAAACG	CTAAAAAGGC	AAATGCACTA	5250
35	CTATCTGCAA	TAAATCCGAG	CAGTACTGCC	GTTTTTTCGC	CCATTTAGTG	5300

GCTATICTIC

5650 5700 5750 5600 5800 5850 5900 5950 6000 6050 6100 6150 6200 6250 6300 6350 6400 6450 6500 6550 9600 6650 6700 6750 6800 6850 6900 TCATAGCTGT CATACGAGCC ATACGGTTAT GCTAACTCAC AACCTGTCGT CGGTTTGCGT GCICGGICGI AAAAGGCCAG TTTCCATAG AGTCAGAGGT CCCTGGAAGC GATACCTGTC TCACGCTGTA CTGTGTGCAC GCAGCCACTG ACTATCGTCT AGAGTTCTTG AGCTCTTGAT TTGCAAGCAG GGTCTGACAG TIGGIAICIG TGATCTTTTC GGGATTTTGG AAATTAAAA TGTCTATTTC TACGATACGG GAGACCCACG GGAAGGGCCG GTCTATTAAT TCGTTTGGTA TTCATCATCA CAATGCGCTG ACTGAGTGTA GTAATCATGG TTCCACACAA ATACCGATCA TAATGAGTGA CCAGTCGGGA CGGGGAGAGG GACTCGCTGC AAAGGCGGTA ACATGTGAGC TTGCTGGCGT TCGACGCTCA TTCTCATAGC TTATCCGGTA AGGCGTTTCC CCGCTTACCG CCAAGCTGGG GCCACTGGCA GCGGTGCTAC AGGACAGTAT AAGAGTTGGT GTTTTTTGT GAAGATCCTT CTCACGTTAA AGATCCTTTT GAGTAAACTT CTCAGCGATC TGTAGATAAC ATGATACCGC CCAGCCAGCC CCTCCATCCA CCAGTTAATA GTCACGCTCG GGCTTGGAAT GATTGGCTAT CATCATGCTA TAAGTGATGT CGCGCTTGGC CTGGGGTGCC CCGCTCACAA CAGCTCACTC GCAGGAAAGA TGCCCGCTTT GGCCAACGCG AAAGGCCGCG CICGCICACI ATCACAAAAA TAAAGATACC TCCGACCCTG GCGTGGCGCT GICGIICGCI CCGCTGCGCC ACGACTTAIC AGGTATGTAG CTACACTAGA CCTTCGGAAA GGTAGCGGTG AGGATCTCAA GGAACGAAAA ATCTTCACCT AAGTATATAT AGGCACCTAT CAGTGCTGCA CTCCCCGTCG AAGTAGTTCG GCATCGTGGT CAGCAATAAA ACTITATCCG CTGCCACAAA AAAAGCCAAC CATCGTTAAA GGGTTAATTG AAATTGTTAT TIGCGCICAC TTAATGAATC CACCGTGCTG AGTGTAAAGC CTTCCGCTTC CGAGCGGTAT AGGGGATAAC GGAACCGTAA CCTGACGAGC GACAGGACTA GCTCTCCTGT TTCGGTGTAG TTCAGCCCGA CCGGTAAGAC TAGCAGAGCG CCTTCGGGAA CTAACTACGG AAGCCAGTTA GCAGAAAAAA AACCACCGCT ATCAAAAAGG TTAATCAGTG GACGCTCAGT AATCAATCTA AGTTGCCTGA CATCTGGCCC CCAGATTTAT TGGTCCTGCA AAGCTAGAGT ATTGCTACAG ACCCGTAATG AATAGCACCA CAACAAATGG CCTTTAGTGA TICCIGIGIG GGAAGCATAA ATTAATTGCG GCCAGCTGCA ATTGGGCGCT CCACAGAATC GCTCCGCCCC GETATCTCAG TGAGTCCAAC rcsscrscss CAAAAGGCCA GGGAAACCC TCCCTCGTGC CGCCTTTCTC SAACCCCCCG GTAACAGGAT AAGTGGTGGC CGCTCTGCTG CCGGCAAACA CAGATTACGC **ICATGAGATT IGAAGITITA** TACGGGGTCT GTTCATCCAT BAGGGCTTAC TTACCAATGC AGCGCAGAAG retreccese CGTTGTTGCC CTCACCGGCT 10 15 20 25 30 35

	IGGCTICALL	CAGCTCCGGT	TCCCAACGAT	IGGCTICAIT CAGCICCGGI ICCCAACGAI CAAGGCGAGI IACAIGAICC /IOO	TACATGATCC	207/
	CCCATGTTGT	GCAAAAAAGC	GGTTAGCTCC	CCCATGITGI GCAAAAAGC GGITAGCICC ITCGGICCIC CGAICGIIGI 7150	CGATCGTTGT	7150
	CAGAAGTAAG	TTGGCCGCAG	TGTTATCACT	CAGAAGTAAG TIGGCCGCAG IGITAICACI CAIGGITAIG GCAGCACIGC 7200	GCAGCACTGC	7200
	ATAATTCTCT	TACTGTCATG	CCATCCGTAA	ATAATTCTCT TACTGTCATG CCATCCGTAA GATGCTTTTC TGTGACTGGT 7250	TGTGACTGGT	7250
٧,	GAGTACTCAA	CCAAGTCATT	CTGAGAATAG	GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG 7300	GACCGAGTTG	7300
)	CTCTTGCCCG	GCGTCAATAC	GGGATAATAC	CICITGCCCG GCGICAAIAC GGGAIAAIAC CGCGCCACAI AGCAGAACII 7350	AGCAGAACTT	7350
	TAAAAGTGCT	CATCATTGGA	AAACGITCIT	TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG 7400	ACTCTCAAGG	7400
	ATCTTACCGC	TGTTGAGATC	CAGITCGAIG	ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA 7450	GTGCACCCAA	7450
	CTGATCTTCA	GCATCTTTTA	CTTTCACCAG	CIGALCITICA GCAICITITA CITICACCAG CGITICIGGG IGAGCAAAA 7500	TGAGCAAAAA	7500
10	CAGGAAGGCA	AAATGCCGCA	AAAAAGGGAA	CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT 7550	ACGGAAATGT	7550
	TGAATACTCA	TACTCTTCCT	TITICAATAT	TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG 7600	TTTATCAGGG	7600
	TTATTGTCTC	ATGAGCGGAT	ACATATTIGA	THATTETTE ALGAGEGGAT ACATATTEGA ALGEATTAG AAAAATAAAC 7650	AAAAATAAAC	7650
	AAATAGGGGT	TCCGCGCACA	AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCAC	AAGTGCCAC		7689

SEQ ID NO:58(a Kozak sequence) ACCATGG 15

SEQ ID NO:59 (a Kozak sequence)

ACCATGT

2

SEQ ID NO:60 (a Kozak sequence) AAGATGT

SEQ ID NO:61 (a Kozak sequence) 25

ACGATGA

SEQ ID NO:62 (a Kozak sequence) AAGATGG

SEQ ID NO:63 (a Kozak sequence) 30

GACATGA

TCTATTCTGGGGGGGGGGGGGGGGGGGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGG GGCGCCTGGATCCAGATCACTTCTGGCTAATAAAAGATCAGAGCTCTAGAGATCTGTGTGTTGGTTTTTTTG **AAGGTGCCACTCCCACTGTCCTTTCCTAATAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAT** ttcaagagtc ggattacata cacccagatt aatgtggett ettegetaet tctgccttag tttatcatca ctttaaaaat aaaaaacaat tactcagtgc ctgttataag cagcaattaa ccttgaatac TGCAAAAGGT AATGCTTATG TGCAGGTTTT tccgggatct cgggaaaagc gttggtgacc aaaggtgcct gatttagatt agtgccaggc aaagcaaaat cctgaggttt taataaaac aaaaaattag aaatggttgg GATGAAGTTG TGAGAATCTC aggtaaacge cattgtcage aaattgatee aagagaacea CTGTAGGAAA AATTGTTCAC ATTTTGCTCC CATTTACTTT GGAAGATAAA ATATTTACAG cgtcactctg aaatgggttt gctgcaggcc aatatttcc attatctagt gatgcctatc attggttgga atgaacttga ctgatttaac tgaacattat cttgattata ttattgataa TGAATGIGIT CITGIGITAI CAATATAAAI CACAGITAGI GCCTTACTGC GTATTTGGTT (avian optimized polyA) SEQ ID NO:69 (vitellogenin promoter) aatccagttc totgccattg ctgcttcctc tgcccttcct ctcaacatct agcaccctt cctgattaac aggcgtttgt ttatgattga tgcctacatc acaacaaaa CAGCTACTTG GCTGCATTTT TATCCCTCTT (a Kozak sequence) (a Kozak seguence) SEQ ID NO:66 (conalbumin polyA) (synthetic polyA) ccatgagaga gaaataaat tteetteece CCTGCACTTC ggggatcgc tctagagcga atcccattgc ctccc SEQ ID NO:64 SEQ ID NO:65 SEQ ID NO:68 ID NO:67 gttaagcaca tgccctagat ctactgctgc gccacagcaa attactggta aaagtatatt tccaagaagt TCTAGGCTGA **IGCATCAGIT** ACCATGT ACCATGA CTCTC S 10 15 20 25 30 35

GATTACGCTC AGTCAATGTA TCAGGTGCCA CCAGCAGCCA GTAAATGCCG ATTIATGATT CAGAAAACAG GCCTCTTAAA AAGGTTTTTT CGGAGCTGAA AGAACACATT CGTGAAGAGA GCCTGGCAGA TTTCTCAGTA GGTTTATCTG GGGAGAGTAT ACAACTTGGG GGACCATGTA TAAATTTATT CAATCAGAAA AGGAATTCAG TCAGCGTGAC TGAGGGAAGA CAGCTAGATA ATCTGTGGCA CTTTATTCGC AAAAAACTTG GTCTCACCAT CCATATATTG GAATAGCAAA AACTTCTGGT ACATCAATTT AAACCAAAGC GGGTTATTAT ATAAACTGAT TTCCATAAAA ATGAATACAT ACCTGGGCAA AACCAGTCTC AGCAAGTAGC GATACATTGA ATTTGATTTT TCATTTAAAA ATATTCCTGG TACATATGTT CTGAACATTC AATCTAGGAA TTTCAATAAA TGGTGCACAT GAATGGCCAT ATTGCTGATT CCAAGGTATT GAGTGTTTAC GITTIACIGI GGTCAACATA GATTATACTG AAACCTTTGT TGGCCTGCAG GCCGTGACCC TAGTAGAAGT ATCAGAGATG GCAAAAAGAG CCACGIGITC GATCCCGTGA TCATGTGCGT CCITCGCI S 10

GAGGICAGAAT GGIIICIIIA CIGIIIGICA AIICIAIIAI IICAAIACAG chicken) AACAATAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG TTGTTCTCAC ACAAGAAGGA AAGCTCAATG AGTCCTGATG GAACACTGCA GCATTGTATG AAAGAAAGTG TICCICIGIC TAAAGCATTC GCATGCTTCT TAATAAGCAC TTCCCACATT TAAAAACTAA TAAAATCTAA SEQ ID NO:70 (fragment of ovalbumin promoter CTATTGAAAC AATTTCACAA CGTTGTAGAT GTIGIAACIG CCCATCCAAC TACCCAGAAT CAAAGGTCAA ATATATTCCC CCATTAAGTT ATTCATGGAA GATCTTTGAG GTATTGTTT AAGTTTGTTA CAGCTATAGG TCCTTCACAT AGGTCACGTC ACAACCATAG ATCTCTAAGG TCTATGGCGT GGTGGGTCAC AATTCAGGCT CCTCCAGCTG ATAAACACAT TTGAAATTGA CTCAGAAAAA GTCTGTTACT TATGTTTCC CAAAATGGAC TTAAATATTC CICAGITGGC AGAAAATAAT CCATTATCTG AGTCTTCTCT AACACATTGT CTCAGATGCA CATGAACACT TIGCIGIAIC AAAGATAGAT ACAGGCAGAA TATTTTGTCA CAGACTTCTG AATATTTCCC CCTAGCATGG AATTGCTAAT CTTAAAGATC CCATTCAATC TTAAATGATT TCACTCTAGT GAACATGAGC GATTAGCAGA TCCCTCGAAC ATCTGCCAGG AGTTCATATC GAGCTATGTT ACACCCATAA ACTAACTAAT TITCIGITIT TATTIGCICT CCCAATCCCA CGICIGCICI TTATTTCTCC TCAAATGAAA TTATGTCCTG 25 30 15 2

SEQ ID NO:71 (chicken ovalbumin ehancer)

35 ccgggctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag

caaaacagct tccattcgac tettteecte acaattaatg caagttctga cttcatatcc agagccttag atgagactgt ttgtactgct ccttccaact cacaatccca aatggttgct caaacagaga ctcacctgga gtctttattc gcaatggcaa aacaaaatct aaagggcagc acaaaactat tcccaggtaa ggggaaacaa accaaggtat tctctccatg caaagaggag gacctgtgag aggtctccga ctggcatagg acaatgtgtg ttgccaccta gcaacaaga aaatgattta tggctacagc attgtagtgg aagaggcttt tataactttc aatgaaatgc gcattgtcaa tcttcaaact ctcacaatat gggcaatcca tctctctaga tacctgattt tgttgcttac ccatcactga gacctgagca taatgtcaaa gtgattggat aaagtgcttt acagacccac aatttccaca cttgggacaa gtgctgggca cttgacctga caqaqaqaa atgaaaaggc gtaggtttta ttttggataa tggtttaggg attcatctgt tgctccttcc gcaagaagat ctgacctttt ttgcacagct Ś 2

GTGGATCAACATACAGCTAGAAAGCTGTATTGCCTTTAGCACTCAAGGCTCAAAAGACAACTCAGAGTTCAC untranslated region) (2, ID NO:72 SEQ ט 15

ACATACAGCTAG AAAGCTGTAT TGCCTTTAGC ACTCAAGCTC AAAAGACAAC TCAGAGTTCA (putative cap site) SEQ ID NO:73

ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAGGA GCTCAAAGTC CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA TCATGTCAGC TCTAGCCATG TCGCTTTGAT AAACGTTCAC TTCGTTCAGC CTTGCAGTGT TGCAGATCAA CAGAAATGTC TTTCAGAGTG TAGAGTGGCA GACAATGAGC TATAATCAAC GATCAAAGTG CATTGTCTTC AATGGCTATG GCACATCTGT ATGATGTTTA ACACAGATAA ATAAGGTTGT TGCCAGAATA TTCAAACAGC ATGGAATTAT AAGCAATGCC TIGCCAGIGG TGGTTAATGC TIGGITIALL AGCTTGAGAG AAGAGAGGAA CATCTGTCTT (Chicken Ovalbumin Signal Sequence) GCTCAGTGTG ACCAAACCAA AGTCAGACAA GAAGACACAC TACCCAATCC CCTATCAACT GCAATGGTTC ATGTACCAGA GAGCTTCCAT GGCCTTGAGC AATGTTATGG TACAACCTCA CAGCACCAGG CAGTATTGAA CAACCAAATC TGAAGAGAGA CTGGGTAGAA AGGCTTGGAA TTCTCAAACT ATTTAAGGAT TGTGCAGATG GGAGGAAAAA GAAGATCCTG TGAAGTCTCA GACCAGITCI GAGACATCCT GACTTTATGC GTGCAAAAGA GATTCGGAGA TGTATAGAGG GCTCCGTGGA TCATCAATTC GGGAGAAAAC AAAGCAAACC CTGAGAAAT TGTTGCCTGA TGACTGAATG GCATGAAGAT ID NO:74 GTATACCTGG AAACTTCCAG CTTGCCAGTA GTGAAGGAAC TCTTCACTTA GCCAGAGAGC CTTCAGCCAA AAAGGACTGT ACTGAGCAAG TACTTACCTC TCAATGGCTT ATGITGGIGC **LTTGAAAAAC** SEQ 2 25 30 35

AGGCAGAGAG GCATCTCCTC AGCAGAGAGC TCAATGAAGC TCTCTGAAGA TATCAAGCAC ATCGCAACCA ACGCCGTTCT CICITCAGCC AATCIGICIG CATGCAGAAA GCTGCAAGCG CCATGCAGCA TGGAGTGGAT GGCATTACTG ACGTGTTTAG CAGCAGAGGC TCCTCTTCTG CTCAAGCTGT CCCCI CTGAAGATAT GTGGTAGGGT GACCATCCAT AGATGTGTTT

5

ID NO:75 (Chicken Ovalbumin Signal Sequence-shortened approx. SEQ ID NO:75 (Chicken Ovalbumin Signal Sequence-successed Arg GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAG

GCTCAAAGTC approx. Sequence-shortened TATTCAAGGA TGTTTTGATG CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA Signal ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT Ovalbumin (Chicken NO:76 A 100bp) SEQ 10

15 SEQ ID NO:77 (vitellogenin targeting sequence)
ATGAGGGGGATCATACTGGCATTAGTGGTCACCTTGTAGGCAGCCAGAAGTTTGACATTGGT

SEQ ID NO:78 (p146 protein)

KYKKALKKLAKLL

20 SEQ ID NO:79 (p146 coding sequence)

AAATACAAAAAGCACTGAAAAAACTGGCAAAACTGCTG

CTTCTACACACCCAAGACCCGGGAAGGCAGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGGCC TTTGTGAACCAACACCTGTGCGGCTCACCTGGTGGAAGCTCTTCTACCTAGTGTGCGGGGAACGAGGCTT CTGGTGCAGGCAGCCTTGCAGCCCTTGGAGGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGC DGTACCAGCATCTGCTCCTCTACCAGCTGGAGAACTCTGCAACTAG SEQ ID NO:80 (pro-insulin sequence) 25

30 SEQ ID NO:81 (spacer)

(GBGG) x

SEQ ID NO:82 (spacer)

35 GFG

SEQ ID NO:83 (spacer) GGGGSGGGGSGGGS SEQ ID NO:84 (spacer) GGGGSGGGSGGGGSGGGS S

SEQ ID NO:85 (repeat domain in TAG spacer sequence) Pro Ala Asp Asp Ala

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala SEQ ID NO:86 (TAG spacer sequence) Pro Ala Asp Asp Ala Pro Ala Asp Asp 10

Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu SEQ ID NO:87 (gp41 epitope) 15

SEQ ID NO:88 (polynucleotide sequence encoding gp41 epitope)

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Lys

20

SEQ ID NO:89 (enterokinase cleavage site)

DDDDK

SEQ ID NO:90 (TAG sequence)

Pro Ala Asp Asp Ala Thr Thr Cys lle Leu Lys Gly Ser Cys Gly Trp 25

Ile Gly Leu Leu Asp Asp Asp Lys

SEQ ID NO:91 (altered transposase Hef forward primer) ATCTCGAGACCATGTGAACTTGATATTTTACATGATTCTCTTTACC 30

atgcccagta tcctttctcg catatcataa tggagttccg attgacgtca atcatatgcc tegetattae actcacgggg cctggagacg ggtgtgggtt gattattgac gtaggcgtgt teegeggeeg ccgcctatag cttggggcct cttcagagac cgcagcgtga cccgcccatt aaaatcaacg tgttgacatt agcccatata cccaacgacc gggactttcc catcaagtgt gcctggcatt gtattagtca tagcggtttg cgtcagatcg ggtggttacg cgatccagcc gacgtaagta atactctgtc acgttgtatc ttttggcacc caaatgggcg ctgtttttgg ttagcctata tttcttccct SEQ ID NO:92 (altered transposase Her reverse primer) catgtccaac attaccgcca attagttcat atggcccgcc tggctgaccg aacgccaata cttggcagta gtacatctac tgggcgtgga tgggagtttg acaccgggac atggtatagc tgacgatact ctatatgcca ctcctttcgc gccattgcat taaatggccc cccattgacg tttagtgaac tgccaagagt gcatgctata 1 ctgacgcgc ctgtagcggc gcattaagcg cggcgggtgt ID NO:97 (pTnMOD (CMV-prepro-HCPro-Lys-CPA) SEQ ID NO:96 (Syn-polyAr-BatE II primer) ttacggggtc ttgacgtcaa ttggctattg cggttttggc agtacatcaa gctataggtg ctagcgcccg ttcccatagt aaactgccca tcaatgacgg ctacttggca acaactccgc gcagageteg tccatagaag ggattccccg tggctcttat ccctattgg tetetattgg (Xho I restriction site) SEQ ID NO:94 (Bcl I restriction site) GATTGATCATTATCATAATTTCCCCAAAGCGTAACC AGAGGICACCGGGICAATICTICAGCACCTGGIA SEQ ID NO:95 (CMVf-NgoM IV primer) atgacgtatg cggcatcaga atggctcttt gccacaacta taqttattaa taqtaatcaa cttacggtaa tgggacttc gtctatataa tgccagcgcc tatattggct cctattgacg ctccaccca aaatgtcgta tgttttgacc attggaacgc tatttacggt cacacccctt TTGCCGGCATCAGATTGGCTAT gacgtcaata ccgctacact cgttacataa atgggtggag catgacctta acggtgggag attgaccatt ccacgttcgc tatqtacatt aagtacgccc catggtgatg ggactttcca atttccaagt ggaacggtgc actctatagg ccatccacgc SEQ ID NO:93 181 241 481 301 361 541 661 ម 421 601 721 781 841 901 CTCGAG TGATCA SEQ 5 10 15 20 25 30 35

aattcacata gatctccacq agcttccaca acttgatatt cttgctccta gccgcacaag tgttgtattc cagtgtagtc caacagctta attgttaggt ctttatctgt gcaaaaacga ctttaaagtg cacggctgac gactaacaga acttggccgt ttatgagaaa agccgacctt aagaggaaaa ctcgacacgg acatgatatg tccaatctca gccatgggtt tatctattcg cggactaggo aatcgccctg ttgggacaag cttaggcatg ggctgcaacc gttttttgtg ataatgatcc gtgccttcct attgcatcgc agcaaggggg ggtacctctc tctctctctc attatttaca ccaccagtgt ccatgtgtga teggeagete attgggctcg gcagctgagt cggtggaggg atagctgaca taaaacgact ctcttaccga catagcgtgg gaatcgaccg atattcgtga ctaaaagcaa ggcatgctag ctgttactct accaatttct gtgatgctgg taagtcgagt tcagcaactt aaaatcagcg cggcaaagga aacttgttaa gtcctgccta tcatgctgct caacagttcg agaaacaagg ggaaattatg acttactcgt aaatgaggaa ggggcagcac ctgtgtgttg ccctcccc gggctctatg attgacccgg ttaacaaatg actatatata agtgcctgtt ggtcccattt ttttattaaa ctcttctccg tcatggtcgc atgcccacca gaagatgcag caccgtcgga gagcgtggag gtgctgttaa accagacata gaattacact aacatcaaac aaaactctca gtatacagtt gactggtctg cacggtcgtt tggaaaccta gttcatgctc ctcattgtca tggtactggt aagaggctga aaaggccgaa tactcagcgt gacttgaaaa gtgctgaaga aaagctcatg acacccaaac cgttttgata aacgtactct acaagggaag ctgttgtttg tacgetttgg ctctagagat gggatgcggt acatcacaac aaaaactgat tttcctaata 9999t9999t tctctcggta acaattactc tecegttgeg attctgcccc tacaggatgg tgcccgcagt cggacatggg ctccagcggc gcacagcaca gtctgaaat agcggcagaa ეენენენენე tttctgcagt acttgactgt aacaaaacat gcgactcgct tgtacttgtt agtcgcacta atgttcaaag agcggaaaac caccacaccg gaagctgggt tttaggctat tattgtataa atctcgctct aaccttccga acctaaaatc tgaaattcga cagctcagag gcttgcgggc cagaaatcga tctattctgg ctacacaata cacacatggt aagatcagag tgccagccat cccactgtcc aggcatgctg tetetetete teggtaceag aaaataaaa ggtcccatgc ccagacttag ccaaagcgag tgcgagcttc tctgtattt tagggtatgt gacttaaggc cagaggtaac tegttgetge ctctttacca tccacaaga tttcagagca taccgagtaa ccgtcccccg gtacqtqttc ccatgggtct gccacgcatt cgatgcccat aatccgttga cagacctagg actcaaagac accacccgtc acttacctgt agattgaaga taacatgttg gccgaacgag tetetetete catcacttta ctaacacagt ggcattctgg aaaatttatt tggctaataa gccttctagt aggtgccact taggtgtcat agacaatagc tetetetete attgatgcct tgacacggac tacaacaacg cgaatctcgg ttacatgatt ccatggtata tccgagccct acagtggagg gcagatggaa tgataagagt tgagcagtac ctgttccttt acgttggctt aacctgccaa aatcgtcacc tegggeaata cttatggtat gegtteeege gcgagcattc actcattgtc gccgtggcgg gtacaatatg tcatctagtc tgccaaattc ctagcaacta aagcgaatgc atgetteaac cacttccagg gaagttttgc ctacgccata ctactagete agatcacttc tgaccctgga attgtctgag aggattggga gatctgctgt tetetetete tetetetete attaattatg tgccttttat 1321 1381 1441 1501 1561 1621 1681 1741 1801 1861 2161 2281 1921 1981 2041 2101 2221 2341 2401 2521 2461 2581 2641 2701 2761 2821 2941 3001 3241 2881 3061 3121 3181 3301 3361 3481

15

S

20

25

30

ggctcgtatg tatggagttc ggcttggggc taggtgtgggg gggatctcca tttggttctg aaccaactta tcccctaatd ccatgattac tccatatcat ttgattattg ccccdccca ccattgacgt gtatcatatg ttatgcccag tgactcacgg ccaaaatcaa cggtaggcgt cgcctggaga cctccgcggc ctaatccata tccttcagag caaattcaca ggagcttcca tecttgetee gtgccgcaca cgcacggctg gttgttgtat ggcagtgtag cagactaaca gataataata aaaaccttat tcccggtaat ccgcggtggc gcttatcgat catcgctatt taccdcctat tattaaaatt atagcccata gatagcggtt tgttttggca gcttagccta tegtgttege cttgaaaaa gatccaagag cactgatgaa tgtcatatga ttatgcttcc tcgatatcaa tagggactt ccgcctggca accgatccag gtgacgtaag tactgttttt cttccatta ttattattta gatcaatcat acagctatga tggagctcca atacgttgta catgttgaca cgcccaacga tacatcaagt acgtattagt cgcaaatggg accgtcagat caatactctg aacatagcgt cggtagcggc gctcggcagc agattgggct aggcagctga aacggtggag taatagctga atatttcgtt caccaccagt gctttacact gaacaaagc ctgcaggaat tggccattgc cctggctgac gtaacgacaa cacttggcag ggtaaatggc aatgggcgtg aatgggagtt gccccattga cgtttagtga ggtgacgata ggctatatgc ggggtcccat gtttttatta ggatattata gctcatggtc ttatattatt ttggaatgaa tcagcaaatt tgaccctgag cacacaggaa acattaccgc tcattagttc cagtacatct agacaccggg cgtgccaaga atgcatgcta tgatggtata caatgcccac atgagcgtgg aagaagatgc cggtgctgtt ccaccagaca gtcaccgtcg tctaaattta atatttttct atacacatct cgtcaatgac cattgacgtc gtgtctgaaa attatcttga ctatcattgg taattctcgt gttaaggtgg ggcacccag ataacaattt atccccggg gattggctat ctcatgtcca aattacgggg aaatggcccg tgttcccata gtaaactgcc tectacttgg gcagtacatc taacaactcc aagcagagct cctccataga geggatteee tttggctctt atgctatagg ctcccctatt tttacaggat cgtgcccgca tccggacatg gcctccagcg aggcacagca gcagcggcag actecegttg 6060606006 cttttctgca tacctatata tttctcaagg aacgccattg tcactaaagg tatctctatt gtgtggtgtt cttagaaagt atatttgaac gaagtgatgc gacggaatgt aatagtaatc aacttacggt taatgacgta tatgggactt tgcggttttg gtetteacee aggtetatat gctgtttga gcattggaac cegetteett ttattgacca actctgtatt gggtacgtgt ggtagggtat cgacgatgaa tggtaaggta aaatcattaa tcactcatta ttgtgagcgg caattaaccc gaactagtgg tegageatea tttatattgg agtatttacg ccctattga caaaatgtcg ggcacaccc ttgccacaac cgccgtcccc ctggtcccat ggccagactt aaqacttaaq gtcagaggta actcgttgct ttccatgggt gtgagttagc ccctatccaa attttggtaa ttgtgtggaa ggccgctcta actagttatt cgcgttacat ttgacgtcaa accatggtga cgggactttc aatacattac aagctttcct gccaagcgcg accgctgacc aatatgtaca caatgggtgg tcaagtacgc tacatgacct ggatttccaa gtacggtggg cgccatccac cgggaacggt agactctata ctatacaccc ttattgacca actgacacgg tatacaacaa cgcgaatctc catccgagcc taacagtgga aggccgtggc tctgagcagt gactgttcct acttcttcgt taatattt acatggctct acgcagatgg tctgataaga 4141 3661 3781 3841 3901 3961 4021 4201 4261 4321 3721 4081 4381 4441 4501 4621 4741 5641 5821 4561 4681 4801 4861 4981 5041 5101 5161 5281 5341 5461 5521 5581 4921 5221 5401 5701

20

30

25

35

15

gctgcaggag agcctctgga ggggctggag tgcacccgtg gataacattt ctctgggggc ggtgtcgtgg gtcctcagga ccagacctac ctcctcagcc tgagcccaaa gggggaccg gacccctgag caactggtac gtacaacagc tggcaaggag catctccaaa ggatgagctg cgacatcgcc tecegtgetg caggtggcag ctacacgcag gcaaatgaal ctatgagctg ggcccctgtg ctgctstgga tgaaggtgac cggaggacc ctcctacccg agggagcacc ctctggctcc cccaccctcc agtggagacc gagcctgacg ggccctttt aacactgatg tatectgtge cccaggtgca cagactatgc cggggattat ctccagggaa cgttatatct tggtcaccgt ccaagagcac aaccggtgac ctgtcctaca gcttgggcac acaagaaagt ctgaactcct tgatctcccg aggtcaagtt gggaggagca actggctgaa tcgagaaac cccatcccg tggacaagag aaaagctttc tctatcccag agaccacgcc ccaggatcac tgcacaacca ctgagagatt aggtgttcag tcataagtga agtcaggcca aggtggaaga tcactctgtt tcaaggcggg tcacgcatga gcagctacct gggaggaag cccccacga ttgctgcagg ccgaaaggta tcccttagag gtccgccagg ggtgggacaa tcaaaaaca tactgtacca tacttccccg tgcccagcac tacaccctgc cagggaaccc acctttccgg ccctccagca accaaggtgg gacaccetca acaaagccgc ctgcaccagg gcaccctcct gaagaccctg ccagccccca gtcaaaggct aacaactaca aagctcaccg catgaggete ccagagccga tccgggatcc agtggggcc tatcataggg ctggtgtgtc ggacaaacgg taccagcaga gccccctcgg agcagccccg tcaccgcgga tetgecette ccagcaccc tacgcggcca agctgccagg ctctcgtgcc atctcaacat tgcgccagag aagagatgat agccgtatat gaattggggc cttccccctg ggtcaaggac cggcgtgcac 8668888338 gatgagctgg caaaattgat taatgccaag cctcaccgtc gacctgcctg cctctacago ggtgaccgtg gcccagcaac atgcccaccg aaaacccaag cgtgagccac caaagccctc accacaggtg gcagccggag ctccgtgatg gggtaaagcg agtgtcccca tgtttattgg caaacgaccc cttgactatc cagcagtggt gcccaaggct caacaacaag caaggccaca gaaggcagat caaaagctac tgcagaatgt gccatgactc tgctgcttcc aaqaaataaa cattccttcc cagtttcggc gaaacgeetg gtattaaaag ccgaggacac gcttggtaaa tcaccatctc tecetecece gcccatcggt tgggctgcct ccctgaccag ttagcaacgt tgaatcacaa aaactcacac tattacacca tggaggtgca tggtcagcgt aggtctccaa tggtggtgga agccccgaga aggtcagcct gctccttctt tcttctcatg cctcggtgtc cagaaaaata agagcaatgg ccctgtctcc atgaggacag actcaactga caatggccac tcctaggtca ttcaagccaa cagtggcctg ccaaacaag gaacttcgc ctgccacage tegttaagea ggaagtccca cagtggccc cctctgccat gctttgtcaa aaaggcagat tgggtcggcc agcctgaaag gggggagtta tccaccaagg atctgcaacg tcttgtgaca tcagtcttcc ttcactttca acagoggooc aactcaggcg ctctacttcc gtcacatgcg acgtaccgtg tacaagtgca gccaaagggc accaagaacc gactccgacg gatgcattgc gtggacggcg gtggagtggg acacagccac caggggaacg aagagcetet agctcaggga tactactgtt gtggtcatct tctgaggagc ttctttgcta ttttcaagag aagctgaccg ggagccgtga accacacct cttgagcagt gtggagaaga gaaggggaa tgcagaggc 1909 6121 6181 6301 6421 6241 6361 6481 6541 6601 6661 6841 6721 6781 6901 1969 7081 7201 7261 7141 7321 7381 7441 7501 7561 7681 7741 7021 7621 7801 7861 7921 7981 8041

20

25

30

35

taaaattaaa ctccgggatc gatgatgatg aaaacctaaa tattgagctt cgtaaaaat ttggtgacga catgaaggtc aacaccggtt ctttgtttgg cttcctgcca caaccaticat ctatcaatgo atcagctttt ctgtttcctg ataaagtgta cgcgcgggga ttatccacaq gccaggaacc gagcatcaca acccaattcg cctgaatggc tgttaaatca aaagaataga aagaacgtgg ttaggggatt ccagccatcc actactatct tcactgcccg ctgcgctcgg taccaqqcqt gtgctttcgc tegtgaetgg agacacgact gaggggttaa gctgatgact aggcaaatgc gcacgaaccc ggggcccggt ataacatcat taatccaaat caccagccaa ccagcaaaag tageteaege tccactatta tggcccacta gaaagcttat atacatgcga cgcggctttt atgggccaat agtggctatt aatgaaaagc gctggattgg atggtcatag agccggaagc tgcgttgcgc aatcggccaa cactgactcg ggtaatacgg ccccctgac actataaaga cctgccgctt caacccggta tcaatatttt ttttacaacg atccccttt agttgcgcag gttaaatttt ttataaatca accaaaatca tcttctttat atgtataccg tgtattatct tgctatttac tgaagctaaa atttcgcgga gctaaaccaa gcaaacagta aaagcacctt gcgatcccac aacgctaaaa tegeceattt cgctttctca cttgcagcac ccttcccaac taaaattcgc gcaaaatccc ggaacaagag aatgatttt gcatatcgca ccctgagctt caagacccgt accacaccgt taaataagtg tggcgtaatc acaacatacg tgcattaatg cttcctcgct actcaaaggc gagcaaaagg ataggeteeg acccgacagg ctgttccgac taggetgtgt gtcttgagtc gaaatccagt ctggccgtcg atcagggcga atteceteag tcacattaat acaggtgttt gagaattaac ctgtttactc ttcagatata tgtaaaagac atcctgatta gceteceete cgcgcgctca acttaatcgc caccgatcgc aatattttgt gccgaaatcg ggctggttat gccaatttat aggttttatt gggtcattat acagtaaaac ataaataaca taatccctgt tgcaggtaaa tgccgtttt ctgtaatagc atggcatcgt attgcgcgct acaattccac gtgagctaac tegtgecage cactetteeg gtatcagctc aagaacatgt gcgtttttcc aggtggcgaa gtgcgctctc ggaagcgtgg cgctccaagc ggtaactatc gcccatgaga gttccagttt accttaactt aaaaccgtct gtgaatgatt ttatccgctc gcgttaccca caaagggcga atgtgtatcc cagggtcaac gttatcaaga ggctcaccaa ctccctgtaa cgagcagtac taatcaacaa gtaggtcgtt cgccttatcc tatgccctag agtcgtatta tgtaagcgtt taaccaatag gttgagtgtt cttactcaat gataaaaag taaaatagat cacttgtctc ggataataat ctaaccaacc gaatactgag atcacgattt gtgagggtta tgcctaatga gggaaacctg gcgtattggg gcggcgagcg taacgcagga cacattacta aagctccctc tetecetteg ttatcccatt aagaggccg ctcaagtcag gctattcatc gctgaaataa gcaccaactt ccctatagtg agagettgee caaattggta gcaataaatc tgtgaaattg gaggcggttt tcagttcggt atggattaca gaaaaccctg cgtaatagcg gctcatttt atatgacaag tgcttaaaaa gaaagataaa gccctcttgg atcgatagca gcattggtaa atagacatca acagggaaa caaaggettg qttcccttta aagcctgggg cttccagtc aatcagggga gtaaaaaggc aaaatcgacg ttccccctgg tgtccgcctt ccgaccgctg ttcacccaca gaatggaaat ccgagatagg actccaacgt catcagtgct aataactaag tegttegget 10501 10141 10261 10381 8701 8761 8941 9001 9061 9121 9181 9241 9301 9361 9481 9541 9601 9661 9841 9901 10081 10201 10321 8401 8461 8521 8581 8641 8821 8881 9421 9721 9781 9961 10021 2 15 25 35

	10561			actggtaaca	ggattagcag	agcgaggtat	gtaggcggtg	
	10621			tggcctaact	acggctacac	tagaaggaca	gtatttggta	
	10681		gctgaagcca	gttaccttcg	gaaaaagagt	tggtagctct	tgatccggca	
¥	10741		cgctggtagc	ggtggtttt		gcagcagatt	acgcgcagaa	
<b>O</b>	10801		tcaagaagat	cctttgatct		gtctgacgct	cagtggaacg	
	10861		ttaagggatt	ttggtcatga		aaggatcttc	acctagatec	
	10921		aaaatgaagt	tttaaatcaa	tctaaagtat	atatgagtaa	acttggtctg	
	10981		atgcttaatc	agtgaggcac	ctatctcagc	gatctgtcta	tttogttcat	
10	11041		ctgactcccc	gtcgtgtaga		acgggaggc	ttaccatctg	
2	11101	-	tgcaatgata	ccgcgagacc	cacgeteace	ggctccagat	ttatcagcaa	
	11161		agccggaagg	gccgagcgca	gaagtggtcc	tgcaacttta	tecgeeteca	
	11221		taattgttgc	cgggaagcta	gagtaagtag	ttcgccagtt	aatagtttgc	
	11281		tgccattgct	acaggcatcg	tggtgtcacg	ctcgtcgttt	ggtatggctt	
1	11341				gagttacatg	atcccccatg	ttgtgcaaaa	
2	11401	aagcggttag			ttgtcagaag (	taagttggcc	gcagtgttat	
	11461			ctgcataatt	ctcttactgt (	catgccatcc	gtaagatgct	
	11521		tggtgagtac	tcaaccaagt	cattctgaga a		cggcgaccga	
	11581			atacgggata a			actttaaaag	
ç	11641		tggaaaacgt	tettegggge ;	gaaactctc a		ccgctgttga	
2	11701	gatccagttc	gatgtaaccc	actcgtgcac	ccaactgatc t		tttactttca	
	11761	ccagcgtttc	tgggtgagca a	_	-	_	ggaataaggg	
	11821	cgacacggaa					0011111000 00001111000	
	11881		-					
1	11941	_		_		rrayaaaaar ,	aaacaaacag	
25					3			
Ŝ	SEQ ID NO:	NO:98 (prnMCS	(CHOVep-prepro-HCPro-CPA))	pro-HCPro-CI	( <b>( V</b> d)			
	-	ctgacdcdcc	ctataacaac	gcattaagcg	Cognoportor	COTTOSTE		
	61				ctcctttccc	TITIOTITION	なかいかいかないかい	
(	121	. ccacgttcgc		ttagctatta	accattocat	acattatato		
30	181			catqtccaac	attaccacca	tattaacatt	gattattgan	
	241		tagtaatcaa	ttacggggtc	attagttcat	agcccatata	tagagttcca	
	301		cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	cccccatt	
	361		atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	
35	421		tatttacggt	aaactgccca	cttggcagta	catcaagtgt	atcatatgcc	
Ç	481		cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	
	547		tgggactttc	ctacttggca		gtattagtca	tegetattae	
	T09		cggrrrggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	
	100		crccaccca	rrgacgreaa	tgggagtttg	ttttggcacc	aaaatcaacg	
	17/	ggacttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt	

aggggatcgc agccgacctt aatcgccctg cttaggcatg ggctgcaacc tttatcatca ttatgattga ggtgtgggtt cttcagagac aattcacata gatctccacg agcttccaca cttgctccta gccgcacaag cacggctgac tgttgtattc cagtgtagtc gactaacaga actcgatatt caacagctta acttggccgt ctttatctgt qcaaaaacga ttatgagaaa ctttaaagtg aagaggaaaa acatgatatg tccaatctca ctcgacacgg gccatgggtt tatctattcg ttgggacaag ccgcctatag aatccataac attgttaggt cqqactaggc cttggggcct cagcaattaa cggcaaagga ggaaattatg teggeagete cggtggaggg atagctgaca ccatgtgcga taaaacgact ctcttaccga gaatcgaccg ggcatgctag atattcgtga gtgatgctgg tcagcaactt ctaaaagcaa aaaatcagcg aacttgttaa qtcctgccta tcatgctgct agaaacaagg acttactcqt ctgtttttgg atactctgtc attgggctcg gcagctgagt ctgttactct taagtcgagt caacagttcg gacgtaagta ttagcctata ttccattact attatttaca catagcgtgg gtagcggcgg ccaccagtgt accaatttct cgtcagatcg cgatccagcc gttcatgctc cgttttgata aacgtactct tacgetttgg ggtcccattt gagcgtggag accagacata caccgtcgga gaattacact aaactctca aacatcaaac gtataccgtt gactggtctg cacggtcgtt aaagctcatg ctcattgtca tggtactggt tggaaaccta aagaggctga aaaggccgaa tactcagcgt acacccaaac gacttgaaaa acaagggaag gttggtgacc ctgttataag gcatgctata tgacgatact ctatatgcca ttttattaaa ctcttctccg atgcccacca gaaga+gcag gtgccgttaa tttagtgaac acaccgggac tgccaagagt atggtatagc tcatggtcgc caqaaatcga tgtacttgtt gaagctgggt agcggaaaac tttaggctat acctaaaatc gcttgcgggc ctacacaata cacacatggt cgggaaaagc agcggcagaa gegacteget agtegeaeta atgttcaaag tgaaattcga aaccttccga cageteagag tactcagtgc cggacatggg gcacagcaca gtctgaaaat tttctgcagt attetgeece acttgactgt aacaaaacat atctcgctct gcagagctcg tccatagaag ggattccccg tggctcttat gctataggtg ccctattgg tctctattgg tgcccgcagt ctccagcggc tecegttgeg ეენენენენე caccacaccg tacaggatgg gccacgcatt cgatgcccat tgcgagette tttcagagca accaccogto acttacctgt ggcattctgg tccgggatct ctttaaaaat aaaaaacaat cagaggtaac tegttgetge ccatgggtct ccaaagcgag taccgagtaa aatccgttga actcaaagac tattgtataa agattgaaga gccgaacgag aaaatttatt gcttccttat gtacgtgttc ccagacttag gacttaaggc ctctttacca tccacaaaga cagacctagg taacatgttg ctaacacagt cacaccctt ggtcccatgc gtctatataa tgttttgacc attggaacgc attgaccact gccacaacta tctgtatttt ccgtcccccg tagggtatgt gcgagcattc ccatggtata tgccaaattc actcattgtc ctagcaacta atqcttcaac gaagttttgc ctactagctc tctagagcga ggaacggtgc tgagcagtac acgttggctt aacctgccaa aatcgtcacc tegggcaata cttatggtat gegtteeege gtacaatatg aagcgaatgc ctacgccata cacttccagg atggctcttt tgacacggac gcagatggaa tgataagagt ctgttccttt ttacacgact tcatctagtc acqqtqqqag attgaccatt tacaacaacg acagtggagg gccgtggcgg ccatccacgc actctatagg atacacccc cgaatctcgg teegageeet 2041 2101 2341 1381 1441 1501 1681 1741 1801 1861 1921 1981 2221 2281 2461 2581 2641 2701 2881 3001 1081 1141 1201 1261 1321 1561 1621 2161 2401 2521 2761 2821 901 1021 961

2

15

10

25

30

tgcctacatc

5

tccaagaagt attactggta ttcctgacgg qqtaaaaatc ttagctcact tggaattgtg gcgcgcaatt ctctagaact tgacctcgag cctgatacct agaaaccatc aaggcaattt tctgtgacct cttcctaatg ttttagtgat aagattgttg tcagttgata tcttaagtcc gataaaaagt tagggacaga cttttcttgg ttcttagtgg cagctgtgct agataacatt aaaggaagga aagtaattct gtacttttt aattcctctg ttttgctgta tcttgttctc gatcaaatga tactgtttgt ttgctattgt caagttcata gctccttcac atcccattat atttaaatat tetgeettag ccttgaatac attacgccaa acttaaagct taatgatttt tatccttttg ctgctttgca caactgcaag cttatcccta gtaatgtgag tegatacege aggagettga cageteagag ttgctlcttt ccctcatgaa taatgtgctc tctgagtagg cttagctgac tcataggtag acaaaatgaa acaaaaaaa gtatgttgtg gtggcggccg tcgacattca actgttggtt tgaatttcac tttcagact taccatcttg tgttttcaca tggtatagga gaattatgtt agaggtcaga atggtttctt tgaaatatat tggagctatg aaaaagatag aggaacactg gtctcagttg ataggtcacg tttcttaaag tatgatcccg cttccggctc tatgaccatg atcaagctta tcacatccaa aaatggttgg taataaaac aaaaaattag aagagaacca atgaatcccc ctccaccgcg tcccacaaa ggcaatccat aaagacaaac agagaacaat ctggacttca gcagcagagc gcaggcaaga aggagaacaa aatctcaagt tattcatgag actatttgta ggtaaccttc gacatgcatt ttatttgcac acagtatata cttctataac ctcctccagc atgtcctgaa cttcagctag aagatctttg ttgcattgta tcacacccat cttcactcta caagaaaata cacgttgtag tagtaataag atttttctgt attggttgga atgaacttga cattgtcagc aaattgatcc ctgatttaac ttattgataa ctgagcactg acactttatg aaagctggag ggaattcgat ccatgaatgg catcttgtca aggaaacagc gactatgaac aacaacaaa gtgagctcac tccgagtett ggtatgcaat ataggaaagg gtgtgacaaa acctatccca tctgcagact gtaattaaaa caaaaaata gtagatttcc atctctatgt atcacaaaaa actgactaaa agtgctttac ttattcatgg agaacaatag accatgaaca tataaagcat cctattttgt gagtattggt tgcgtctgct ggctcagatg cctacaacca aggttcccat ctcgttgacc ggtggataca aaacagggga aatatgcgac acaacaaaa cttgattata cccaggcttt aatttcacac aaagggaaca ccgggctgca atgccaggtg atttctctct agtggcaaag ctttcaggtc gctttgacct aatgcctggc ctagaaagct cacatctatc acagcaccaa gtcaaacaat atccattgcc agcaaaagac gattgcagtg agacatccag agaaaatggc atttcaatac taaatcaaac tgtccctcga atttgaaatt tggtctgtta caaggagaat ggccattaag ggaaagaaag ctttatttct tgcctagcat aaaagtttgt tccatctcta aatgttaatt tqaacattat gatgcctatc aggtaaacgc attaagttaa agcggataac ctgcagaaa gattttcttc actgatggct cattaggcac aaccctcact agtggatccc gagcaaaatq ccacactcac tcaaaattgt tggataagag gggcaggcg gcttttataa cccacaatga gacaagcatt tactgggaag cttactctct ctgaaataga cctgtggtgt tcagacttag gcttaacaga ccccatttt caattctatt atattatgat tcataaacac tcctcagaaa gagaaacaaa tcatctgcca tccagctata atgcatgctt acttatgtcc aacagacttc aattatgttt 3181 3301 3361 3421 3241 3481 3541 3601 3661 3841 3721 3781 3901 3961 4021 4081 4141 4201 4261 4321 4381 4441 4501 4561 1621 4681 4741 4801 1861 4921 4981 5041 5101 5161 5221 5281

15

25

30

tctgggggca gtgtcgtgga tcctcaggac cagacctaca gagcccaaat gggggaccgt accettgagg aactggtacg tacaacagca ggcaaygagt atctccaaag gatgagctga gacatcgccg cccqtgctgg aggtggcagc tacacgcaga gagetgacae tctggagatg cctgtggtgg ggtgactact gggaccaagc ccctcctctg gagaccacca gcacccgtga caaatgaata ataacatttg tecteageet ggctccagct aattaaaaac catctcgtga ttggttctgg gcctctggat attaaaattt ctgcaggagt gggctggagt cccagtcttc ggtcaagttc ggacaagagc gcacaaccac gatcacctgc aggccaggcc gttcagcgga ggcgggagtg gttatatctg ggggattatg ggaggagcag cgagaaaacc cccatcccgg ctatcccago gaccacgcct gctttcctat gagattetet ggaagatgaa tctgttccca aagtgactcc tatttcgttt cgtgttcgct ccaggtgcag ggtcaccgtc accggtgacg tgtcctacag tgaactcctg tgttacccag atcaatcatt ctcctgtgca tccagggaag caagagcacc cttgggcacc caagaaagtt gatetecegg ctggctgaat aactaaaatc aagggaacct agactatgct gccccgtcaa gtgggacaac caccetecte acttccccga aagaccctga acaccctgcc agctcaccgt cagagccgaa aaacggccag agcagaagtc ggatccctga cctcggtcac tgtgtctcat agcaatattt tatttttctt cgaaaggtac actgtaccac cctttccggc ccaaggtgga caaagccgcg tgcaccagga cagcccccat tcaaaggett acaactacaa atgaggetet gggcccaggt atagggaggt gacctattga caaacttctg cagccagtgg ctaaatttaa cccttagagt taagaaaga caaaaacac agggancect cctccagcag gcccagcacc acaccctcat gaaaacacat agtggttatc gtgaccgtgc ggtaaagcgc tccccaggac cgaccctccg gccacactgg gcagatagca ttctcaagga gcgccagagc gccgtatatt aattggggcc ttccccctgg gtgagccacg aatgccaaga cagccggaga ctctacagca tecgtgatge tattggtacc actatcagtg aaggctgccc atggaacatg atccaaaatg acctatatat atgagctggg aaaattgatg agagatgatt gtcaaggact ggcgtgcaca cccagcaaca tgcccaccgt aaacccaagg ctcaccgtcc aaagccctcc ccacaggtgt acctgcctgg aqaacaqqca tgtcaaaggt ccccagggct ccgggggggt ggtctccaac ggtggtggac ggaggtgcat ggtcagcgtc gccccgagaa ggacagcaaa ggccaccttg aggtcagccc agccaacaag ggcctggaag caccatctca cttccccca ggtcagcctg cttctcatgc aactgacagc tgtggtgtt gacgatgaat agtttcggct tattaaaagc cccatcggtc gggctgcctg tagcaacgtg aactcacaca gagcaatggg ctccttcttc ggtgtcagtg aaaatatgtt ctgaagctca atggattagc tetecattea atttctatgg gctatatatt cttggtaaag aaacgcctgg cgaggacaca ccctcccccg cctgaccagc gaatcacaag cctgtctccg actgttactc aggagettea ccgtgacagt ggggagttat agagcetete aatatatttc ctttgtcaac ccaaagggca actccgacgg agccacctc cattgccaga tcatctatga tgaccgtcct ctggttgtaa ccattaaatg cttcttcgtg cggggggagg tcactttcag gggtcggccg aaggcagatt gcctgaaagc ccaccaaggg actcaggcgc cttgtgacaa tcacatgcgt acaagtgcaa tggagtggga aggggaacgt cagggacaat aacagtcctg taatatttgc cacaattcag cagoggcoct tctacttcct tctgcaacgt caqtcttcct tggacggcgt cgtaccgtgt ccaagaacca 6301 6541 6601 6781 6841 7141 7261 7501 5521 5581 5701 5761 5821 5881 5941 6001 1909 6121 6181 6241 6361 6421 6481 6661 6721 1069 6961 7021 7081 7201 7321 7381 7441 7561 7621

15

2

20

25

30

cacctccaa

agcaccgtgg cttttgaag ctgatgtgca tgtggcttct tgagatttt tgccaggcac agcaaaatgg tttcgcttca cgaatggaaa gcgtaatagc accgagatag catatgacaa tcatcagtgc gtgcttaaaa tgaaagataa ggaaaaccct ttgttaaatc agctcatttt gactccaacg tgccctcttg gatagacatc gccctatagi aagagcttgc aaataactaa categrage ccaaattggt tgcattggta aacagggaaa tgcaataaat aaagcctggg acaaaggctt tgctattcat cqctqaaata tgttcccttt tgtgaaatt gctttccagt gaatcagggg ctacctgage tcactctgaa ttatctgtgc ggcaaaatcc cttacaaatc aaaagaatag gcatgaaggg gggaaggcc gcacggaaca tgcaggccag tacccaattc tegecagetg gcctgaatgg tggaacaaga gtccactatt aaagaacgtg taccaaaatc attaggggat tttggtgacg atgggtttcc tattttccaa gtcgtgactg actccgggat tgatgatgat aaaaacctaa ttattgagct tcgtaaaaa cactactatc acatgaaggt tccagccatc taacaccggt tgctgatgac tctttgtttg ccaccagcca ataaaattaa gctatcaatg tetteetace gatcagcttt ccaaccatca cataaagtgt ctcactgccc gctgtttcct gttatccaca cggccagcag cccttcctcg gttttacaac cccttcccaa cagttgcgca gccaggtcac cgcggaggga cgtgcccccc caacatctaa cacceettge tccaqttcaa gtgttttgta 8999966699 catccccctt cgttarattt atggcccact ggaaagctta aatacatgcg ccgcggcttt atcttcttta aataacatca tgaggggtta ataatccaaa aatgggccaa tagtggctat aaggcaaatg taatgaaaag tgctggattg gatgtatacc catggtcata gagccggaag ttgcgttgcg gaatcggcca tcactgactc cggtaatacg actggccgtc acaaagcaac aacaagtacg gaatgttcac getteeteg tgattaacag ccttgcagca agctacagct tgactcctct aataaaatct ccttccccag atgagagaa cccctcgagg ttaaaattcg tatcagggcg taatgattt tgcatatcgc ttgctattta ttgaagctaa tatttcgcgg cgctaaacca cattccgtca taaagcacct aaacgctaaa cccctgagct agcaaacagt agcgatccca ttcgcccatt ttaaataagt cactcaaagg ccaagacccq caccacaccg ttggcgtaat cacaacatac ctcacattaa ctgcattaat getteetege cttcgcgcca gtcccacaaa ggcccctgca tgccattgct cacagcaaga ccctagatcc aacttaatcg taagcacatt actgctgccc cccattgcct acgcgcgctc gcaccgatcg ggccgaaatc tgttccagtt aaaaaccgtc caccttaact tggctggtta ggccaattta taatattttg cgagaattaa taggttttat agggtcatta cctgtttact tacagtaaaa tataaataac cttcagatat ataatccctg atgcaggtaa ctgccgtttt tctgtaatag gtcgtgccag gtgtaaaaga aatggcatcg aattgcgcgc cacaattcca agtgagctaa gegetettee ggtatcagct gaggacata ttgctactgc caacttggct ggcgttaccc agcagtggaa aqaaqacaqt ggggaggaaa caagagtcgt attacatatq cccacattat gagtcgtatt ttgtaagcgt ggttgagtgt ggttatcaag agtgaatgat gaagagccc ttaaccaata tcaaagggcg gatgtgtatc tcagggtcaa acttactcaa cgataaaaa ataaaataga gcacttgtct aggataataa aggeteacea actccctgta actaaccaac ccgagcagta catcacgatt gtgcctaatg tgcgtattgg ggaatactga ataatcaaca agtgagggtt gttatccgct tgcggcgagc cgggaaacct 8041 7981 8101 8161 8221 8281 8341 8401 8461 8641 8521 8581 8701 8761 8821 8941 9001 8881 9121 9241 9181 9301 1906 9361 9421 9481 9661 9721 9901

30

attgacgtca atgcccagta aaaatcaacq tccgcggccg cttggggcct atcatatgcc togctattac actcacgggg cctggagacg ccgcctatag ggtgtgggtt cttcagagac aatccataac aattcacata gatctccacg agcttccaca cttgctccta gccgcacaag cacggctgac tgttgtattc cagtgtagtc gactaacaga actegatatt caacagctta acttggccgt attgttaggt gcaaaacga ttatgagaaa gtaggcgtg ctttatctgt agccgacctt ctttaaagtg tatctattcg aagaggaaaa acatgatatg tccaatctca ctcgacacgg gggactttcc ttttggcacc gcctggcatt gtattagtca tagcggtttg gacgtaagta ctgtttttgg ttagcctata catcaagtgt caaatgggcg cgtcagatcg cgatccagcc ttccattact atactctgtc attatttaca catagogtgg gtagcggcgg teggeagete ccaccagtgt attgggctcg cggtggaggg gcagctgagt atagctgaca ccatgtgcga taaaacgact gaatcgaccg atattcgtga ctcttaccga ggcatgctag accaatttct aaaatcagcg ctgttactct gtgatgctgg taagtcgagt tcagcaactt ctaaaagcaa cggcaaagga tgaaattcga acacccaaac aacttgttaa tttagtgaac acaccgggac tgccaagagt tggctgaccg aacqccaata cttggcagta gtacatctac tgggcgtgga tgggagtttg cccattgacg atggtatagc taaatggccc ctatacgcca ggtcccattt gcatgctata tgacgatact tcatggtcgc gagcgtggag ttttattaaa ctcttctccg atgcccacca gaagatgcag gtgctgttaa accagacata caccgtcgga aaaactctca aacatcaaac gactggtctg aaagctcatg gaattacact gtataccgtt cacggtcgtt tggaaaccta aaaggccgaa tggtactggt aagaggctga tactcagcgt ctcattgtca ctacttggca atggcccgcc ttcccatagt tcaatgacgg ttgacgtcaa acaactccgc aaactgccca agtacatcaa gcagageteg tccatagaag ggattccccg tggctcttat ccctattgg tctctattgg cggacatggg gctataggtg tacaggatgg tgcccgcagt ctccagcggc gcacagcaca gtctgaaaat tecegttgeg იმიმიმინი tttctgcagt agcggcagaa attetgeece acttgactgt gcgactcgct atgttcaaag acctaaaatc aacaaaacat tgtacttgtt agtcgcacta caccacaccg gaagctgggt agcggaaaac tttaggctat tattgtataa atctcgctct cttacggtaa atgacgtatg tgggactttc cggttttggc gtctatataa tgttttgacc attggaacgc tatttacggt cctattgacg ctccaccca aaatgtcgta cacaccctt gcttccttat attgaccact gccacaacta tctgtattt gtacgtgttc ggtcccatgc gacttaagge cagaggtaac tegttgetge ccatgggtct acgtacaag ccagacttag tagggtatgt ctctttacca gccacgcatt tttcagagca actcaaagac accaccotto tccacaaaga cgatgcccat tgcgagette taccgagtaa ccaaagcgag aatccgttga cagacctagg acttacctgt cgttacataa ggaacggtgc gacgtcaata aagtacgccc catggtgatg ggactttcca ccatccacgc actctatagg atacacccc tgacacggac atgggtggag catgacetta atttccaagt atggctcttt tacaacaacg acggtgggag attgaccatt cgaatctcgg teegageeet acagtggagg gccgtggcgg gcagatggaa tgagcagtac aatcgtcacc tgataagagt ctgttccttt acgttggctt aacctgccaa tegggcaata actcattgtc ttacacgact gegtteeege gcgagcattc ccatggtata tgccaaattc ctagcaacta cttatggtat qtacaatatg tcatctagtc 661 721 1141 1261 781 1021 1081 1321 1381 1441 841 1201 1501 1561 901 1621 1681 1741 1801 1861 1921 1981 2041 2101 2161 2221 2281 2341 2401 2461

2

25

30

35

ggtaaaaatc tggaattgtg gcgcgcaatt ctctagaact tgacctcgag agaaaccatc aaggcaattt ttttagtgat gataaaagt aagattgttg ggctgcaacc aggggatcgc attactggta cctgatacct cttcctaatg tagggacaga ctttcttgg cagctgtgct agataacatt ttcttagtgg tcagttgata aaaggaagga tgaatttcac aattcctctg cggactaggc aatcgccctg cttaggcatg tttatcatca ttatgattga aaagtatatt tccaagaagt ttcctgacgg ttagctcact tctgtgacct tcttaaqtcc aagtaattct tactgtttgt tggtatagga tgaaatatat taatgtgctc cttagctgac tctgagtagg ctgctttgca taccatcttg gaattatgtt ggaaattatg aaaggtgcct cagcaattaa tetgeettag ccttgaatac taatgatttt gtaatgtgag gtatgttgtg attacgccaa gtggcggccg tegatacege gactatgaac tcacatccaa aggagcttga cageteagag tegacattea ccctcatgaa tatccttttg actgttggtt caactgcaag tcataggtag ttttcagact acaaaaaaa tgttttcaca tcatgctgct caacagttcg acttactcqt cttatcccta acttaaagct agaaacaagg aagagaacca ggtaaccttc gacatgcatt acaaaatgaa agaggtcaga cttctataac ctcctccagc taataaaac ggcaatccat agagaacaat ctggacttca actatttgta gcaggcaaga ttatttgcac acagtatata cttcagctag atgtcctgaa atgaatcccc atcaagctta tcccacaaaa aatctcaagt tattcatgag acaagggaag tacgetttgg gttggtgacc ctgttataag aaatggttgg tatgatrccg tatgaccatg ttgcttcttt gcagcagagc gacttgaaaa cgttttgata gttcatgctc aacgtactct aaaaaattag cttccggctc ctccaccgcg tactcagtgc ctgatttaac ttattgataa atgaacttga aaattgatcc aggaaacagc aaagctggag aacaacacaa ccatgaatgg aaagacaaac aggagaacaa gtgagctcac tccgagtctt ataggaaagg acctatccca atcacaaaaa gtaattaaaa caaaaaata gtagatttcc actgactaaa atctctatgt agtgctttac agaacaatag accatgaaca ctgagcactg catcttgtca acactttatg ggaattcgat ggtatgcaat gtgtgacaaa ctacacaata tctgcagact aaccttccda gcttgcgggc cagaaatcga cacacatggt cageteagag cgggaaagc aaacagggga aatatgcgac aatgcctggc gattgcagtg tgtccctcga attggttgga cattgtcagc ggtggataca cccaggettt aatttcacac aaagggaaca ccgggctgca atgccaggtg agtggcaaag gctttgacct cttcaggtc gtcaaacaat atccattgcc cacatctatc agcaaaagac agacatccag caaggagaat agaaaatggc taaatcaaac atttcaatac ggcattctgg tccgggatct aaaaaacaat acaacaaaa cttgattata ctcgttgacc acagcaccaa atttctctct ctagaaagct agattgaaga taacatgttg ctaacacagt aaaatttatt gccgaacgag aagcgaatgc aatgttaatt agcggataac ctactagctc gatgcctatc aggtaaacgc cattaggcac agtggatcc gattttcttc gagcaaaatg ccacactcac tggataagag cccacaatga gacaagcatt tactgggaag gcttaacaga caattctatt atattatgat tctagagcga ctttaaaaat tgcctacatc tgaacattat attaagttaa aaccctcact ctgcagaaaa actgatggct tcaaaattgt gcttttataa gggcagggcg cttactctct ctgaaataga cctgtggtgt tcagacttag gagaaacaaa ccccatttt ctacgccata atgetteaac cacttccagg gaagttttgc 3121 3181 3241 3301 3361 3421 3481 3541 3661 3961 4081 4141 4201 4261 4501 4621 4741 2881 2941 3001 3061 3601 3721 3781 3841 3901 4021 4321 4381 4441 4561 4681 4801 4861

25

30

35

15

ttttgctgta atttaaatat gctccttcac tottgttoto gatcaaatga atcccattat aattaaaaac aataattgct taacccaatc tctcccatcc catctcgtga ttggttctgg gcctctggat gggctggagt gcacccgtga caaatgaata ataacatttg tecteageet tctgggggca gtgtcgtgga tecteaggae cagacctaca attaaaattt ctgcaggagt gagcccaaat 9599959991 ggggaccgt tacaacagca atctccaaag accctgagg aactggtacg ggcaaggagt gatgagctga aggtggcagc gacategeeg cccgtgctgg tatgagctga aggaacactg gtctcagttg tttcttaaag tggagctatg aaaaagatag ataggtcacg atacaagaag cacactaact cccagtcttc tgttacccag aactaaaatc aagggaacct atcaatcatt tatttcgttt cgtgttcgct ccaggtgcag ctcctgtgca tccagggaag gttatatctg ggggattatg agactatgct ggtcaccgtc caagagcacc tgtcctacag accggtgacg cttgggcacc caagaaagtt ggtcaagttc tgaactcctg gatctcccgg ggaggagcag ctggctgaat cgagaaaacc cccatcccgg ctatcccagc gaccacgcct ggacaagagc gcacaaccac aaagctttcc aagatctttg ttgcattgta tcacacccat cttcactcta cacgttgtag tagtaataag caagaaata atttttctgt tccgccaggc agcaatatt gacctattga caaacttctg cccttagagt gaaacacat cagccagtgg ctaaatttaa tatttittt cgaaaggtac gtgggacaac caaaaaacac actgtaccac agggaaccct caccttcttc acttccccga cctttccggc cctccagcag ccaaggtgga gcccagcacc acaccctcat tgcaccagga aagaccctga caaagccgcg cagcccccat acaccctgcc acaactacaa ageteacegt cagagccgaa tcaaaggett atgaggetet ttattcatgg cctattttgt tataaagcat aggttcccat atccaaaatg tgtcaaaggt ttctcaagga aattggggcc gagtattggt tgcgtctgct ggctcagatg cctacaacca atggaacatg agaacaggca ccccagggct acctatatat gegecagage cc99999999t atgagctggg aaaattgatg gccgtatatt aaacccaagg agagatgatt ttccccctgg gtcaaggact ggcgtgcaca gtgaccgtgc cccagcaaca aatgccaaga ctctacagca tccgtgatgc tgcccaccgt gtgagccacg ctcaccgtcc aaagccctcc ccacaggtgt cagccggaga acctgcctgg ggtaaagcgc ggccattaag atggattagc atttgaaatt aaaagtttgt ggaaagaaag ctttatttct tgcctagcat tggtctgtta tccatctcta ctgaagctca tctccattca atttctatgg gctatatatt tgtggtgtt gacgatgaat agtttcggct cttggtaag tattaaaagc caccatctca cccatcggtc tagcaacgtg ggtctccaac aaacgcctgg cgaggacaca ccctccccg gggctgcctg cttcccccca ggtggtggac ggaggtgcat ggtcagcgtc cctgaccagc gaatcacaag aactcacaca gccccgagaa gagcaatggg ctccttcttc cttctcatgc cctgtctccg ggtcagcctg tcataaacac atgcatgctt tcatctqcca acttatgtcc tcctcagaaa tccagctata aacagacttc ctggttgtaa aacagtcctg ccattaaatg aattatgttt taatatttgc cacaattcag cttcttcgtg aatatattc ctttgtcaac C999999999 tcactttcag gggtcggccg aaggcagatt gcctgaaagc ggggagttat actcaggcgc ccaccaaggg tctacttcct tctgcaacgt cttgtgacaa cagoggccct tcacatgcgt tggacggcgt actccgacgg cagtetteet cgtaccgtgt acaagtgcaa ccaaagggca ccaagaacca tggagtggga aggggaacgt agagcetete 5101 5161 5401 5221 5341 5521 5581 5641 5461 5881 5941 5701 5761 5821 6001 1909 6121 6181 6781 6241 6301 6361 6421 6481 6541 6601 1999 6721 6901 7021 7081 7141 7261 6841 6961

20

25

30

35

15

tgctttcgct gtggactcca gattcatcag gtggagacca gccctttttg tcctgagatt ctgggaaaac ctggcgtaat tggcgaatgg atcagctcat tagaccgaga gatcatatga gatgtgctta gcttgaaaga tgctctggag gccctgtgg ccaccctcct acactgatgt gaatgtggct cagtgccagg caaagcaaaa ttcgccctat taaaagagct aaatgccctc acgaaataac ggtcatcgat atcccaaatt ggttgcattg ttggatagac taaaacaggg tctggctcca gaaggtgact ggagggacca tcctacccgg agcctgacgc gggagcaccc tcatgctatt ggtgttcagc attaaagaac caggatcacc ggtggaagat gtattatctg tttttgttaa atcaaaagaa atcattaggg tcatttggtg aaatccagcc gtcaggccag tgagagattc cactctgttc cataagtgac caaggcggga cagctacctg cacgcatgaa ggagggaagg cccgcacgga tegteactet taaatgggtt tgctgcaggc caatattttc cggtacccaa aacgtcgtga ctttcgccag gcagcctgaa actactccgg ttatgatgat gcgaaaaacc tttttattga ttatcqtaaa ttaacatgaa caataacacc gactettgt ccaataaat atgcactact tattcttcct aagccaacca gtcgttttac atttagtggc aggggggcc gcacatcccc caacagttgc tegegttaaa tcccttataa gcgatggccc ttttaccaaa tcaggaaagc agtaatgggc aaaaaggcaa ccgtaatgaa cccctcggt tggtgtgtct acgcggccag ctgcccttcc tctcaacatc aaatccagtt agagtccact ttaccgcggc taaatcttct cggaataaca ccaataatcc accagcagaa ccgggatccc gtggggcca caccgcggag tctcgtgccc caggtgtttt cqcaatacat gcttgagggg ccttgctgat ccaccaccag atcataggga gcagccccgt gctgccaggt cagcauccct ccatgactcc gctgcttcct agaaataaaa cccatgagag ctcactggcc tegeettgea tegecettee atcggcaaaa gtttggaaca taacattccg tatttgaagc aacagcaaac ctgtaaagca ttttcgccc agaccaagac gtgtccccag agcagtggtt cccaaggctg aaggccacac aaggcagata aaaaqctaca gcagaatgtt atteetteee tcctgattaa ttgttaaaat acttaatgat ttatgcatat ttattgctat ttatatttcg aaacqctaaa taaagcgatc tataaacgct ttgactatca aacaacaagt gtctatcagg acteceetga gtttattggt aaacgaccct cctcccctcg ctcaactgac gaagtcccac tgccacagca cgttaagcac tgttgttcca gcgaaaaacc atccacctta caatggctgg agataggttt aagagggtca tctcctgttt taatacagta gattataaat ccaataatcc gtaatgcagg tgagtgtaaa aatggccacc agtggcctgg agtggccct aaacttcgcg gctactgctg atgccctaga attacgcgcg cccgcaccga ataggccgaa caacgagaat aaccttcaga gtactgccgt ctcggtgtca agaaaaatat tgaggacagc cctaggtcag tcaagccaac caaacaagc ctctgccatt tatcccattg cccaacttaa cgttaatatt aaaggccaat cttggaatac cacaqccacc gctcagggac actactgtta gagccgtgac ttgggttatc agcaggataa ggtagtgaat aaaactaacc atgcattgcc tggtcatcta ctgaggagct ccacaccctc ttgagcagtg tggagaagac aagggggagg gcagaggcc tettigetae tttcaagagt caccaacttg tggattacat tcacccacat agtgagtcgt cctggcgtta agcgaagagg aaattgtaag tttttaacca tagggttgag acgtcaaagg caagatgtgt aaaacttact tgccgataaa taaataaat taagcacttg gtaaggctca aatccgagca agetgacegt tgctcagggt atcactccct 7741 8041 8101 8281 8401 8581 8641 8701 8881 8941 9001 9181 9301 7621 7681 7801 7861 7921 7981 8161 8221 8341 8461 8521 8761 8821 9061 9121 9241 9361 9421 9481

20

30

25

35

15

tttgttccc cctgtgtgaa tgtaaagcct teggtegtte cccctttcc gggagaggcg acagaatcag aaccgtaaaa cacaaaaatc gegttteece tacctgtccg tatctcagtt cagcccgacc gacttatcgc ggtgctacag ggtatctgcg ggcaaacaaa agaaaaaag atccttttaa tctggccca aacgaaaact tctgacagtt tcatccatag ttgcgcaacg gcttcattca gcaatdaacc tccatccagt aaaaagcgg ttgagatcca ataggggttc ttatcactca tgattttatg ccgagttgct aaagtgctca ttcaccagcg agggcgacac acggttatcc accgatcago gcgctcactg aagcataaag ccaacgcgcg ctcgctgcgc aaaggccagg tgacgagcat aagataccag gcttaccgga acgctgtagg accccccgtt ggtaagacac gtatgtaggc gacagtattt ctcttgatcc gattacgcgc cgctcagtgg cttcacctag gtaaacttgg tctatttcgt gggcttacca catgttgtgc agatttatca tttatccgcc agttaatagt gtttggtatg ggccgcagtg atccgtaaga tatgcggcga cagaacttta cttaccacta aaagggaata ttgaagcatt aaataaacaa atcttttact aatcatggtc tacgagccgg agtgatgtat taattgcgtt aatgaatcgg aaggccagca teegeeeeee caggactata cgctcactga aggcggtaat cgaccctgcc ctcatagctc gtgtgcacga gcagagcgag acactagaag agtccaaccc gagttggtag gcaagcagca cggggtctga caaaaaggat gtatatga cagcgatctg cgatacggga caccggctcc gtcctgcaac gtagttcgcc cacgetegte catgatccc gagaatagtg gtatttagaa gaagtaagtt ctgtcatgcc cgccacatag gatcttcagc atgccgcaaa ttcaatatta tctcaaggat tagcaccaca tegttaaata cgcttggcgt ccacacaca cagctgcatt atgtgagcaa ttccataggc tetectgtte gtggcgcttt taactcacat gctcactcaa cgaaacccga aagctgggct tatogtettg ttcggaaaaa tacgattact aacaggatta aactacggct ttttttgttt atcttttcta gattttggtc atgagattat tagataacta cgcagaagtg gctagagtaa tcaatctaaa gcacctatct gacccacgct tgctacaggc atcgtggtgt ccaacgatca aggcgagtta atcgttgtca aattctctta gggcgaaac gcacccaact ggaaggcaaa atatttgaat aagtcattct gataataccg gtgccac acaaatggca atgagtgagc actgtagtga atttctqtaa agcggtatca gctggcgttt atccggtaac gttaattgcg gctcacaatt tgggcgctct aggaaagaac tcagaggtgg cctcgtgcgc ttcgggaagc cactggcage agecactggt gtggtggcct gccagttacc tagcggtggt agatcctttg aagttttaaa cgttcgctcc gataccgcga aatcagtgag ccccgtcgtg aagggccgag ttgccgggaa cggtcctccg agcactgcat gtactcaacc gtcaatacgg acgttcttcg agcaaaaca gageggatae tccccgaaa aatactcata acceactegt catcatcacg tttagtgagg ggggtgccta gtttgcgtat aggccgcgtt ataataatca attgttatcc agtcgggaaa ggctgcggcg gggataacgc ctggaagctc cetteece gacgeteaag cggtgtaggt gctgcgcctt agttcttgaa ctctgctgaa ccaccgctgg cacgttaagg attaaaaatg ttgcctgact gtgctgcaat gatctcaaga accaatgett agccagccgg ctattaattg gatacggtta tggttatggc ttgttgccat tgactggtga cttgcccggc tcattggaaa tttctgggtg ttagctcctt gttcgatgta ggaaatgttg cgcgcacatt attgtctcat 9781 9841 9901 10261 10381 19501 0021 10081 10141 10201 10321 .0441 .0501 10621 10681 10801 10741 10861 10921 10981 11041 11281 11101 11401 11161 11221 11341 11461 11581 11641 11761 11881 11521 11701 11821 11941

25

30

35

15

:100 (pTnMCS rtgacgcgcc ct ccgctacact ccacgttcgc tatgtacatt	egttacataa ctracygtaa arggecegee 1990-1990 gacgtcaata gggacttCo argggtcgagg tatttacggt aaactgcca cttggcagta catcaagtgt aagtacgccc cttggcagta catcaagtgt aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt catgacctta tgggactttc ctacttggca gtacatctac gtattagtca catgacctta tgggactttc ctacttggca gtacatctac gtattagtca	atticogat cicaccca tigacgicaa ggacticaag cicaccca tigacgicag acggiggag gictatataa gcagagcicg calccacge tgittigace tecatagaag	actetatagg cacaccectt tggetettat geatgetata etgtttttgg atacaccec getteettat getataggtg atggtatage ttagectata attgaccatt attgaccact ecectattgg tgacgatact ttecattact atggetettt gecacaacta tetetattgg etatatgeca atactetgte tgacacggae tetgtatttt tacaggatgg ggteccattt attattaca	tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa cgaatctcgg gtacgtgttc cggacatggg ctcttctccg tccgagccct ggtcccatgc ctccagcggc tcatggtcgc acagtggagg ccagacttag gcacagcaca atgcccacca gccgtggcgg tagggtatgt gtctgaaaat gagcgtggag gcagatggaa gacttaaggc agcggcagaa gaagatgcag	tgataagagt cagaggtaac tcccgttgcg grgctgrtaa cygtyyayyy tgagcagtac cagaggtaac tcgttgctgc cacagacata atagctgaca ctgttccttt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga ttacacgact ctctttacca attctgcccc gaattacact taaaacgact acgttggctt gccacgcatt acttgactgt aaaactctca ctcttaccga aacaaaacat aacatcacac caatcaccga aacacaaacat aacatcacaa cgaatcgacga	aatrgtracc tccacaaga tcgggcaata cgatgccat cttatggtat tgcgagcttc gcgttcccgc tttcagagca
NO:100 (p) 1 ctgacgcg 61 ccgctac 121 ccacgtl 181 tatgtac 241 tagttac	301 cgccaco 361 gacgtco 421 atgggtco 481 aagtaco 541 catgaco	-		_ 0 _ 10 0.0.	1621 tgataa 1681 tgagca 1741 ctgtto 1801 ttacao 1861 acgttg	
SEQ ID	10	15	20		30	

cggactaggc ctttaaagtg aagaggaaaa acatgatatg tccaatctca ctcgacacgg gccatgggtt tatctattcg aatcgccctg ttgggacaag cttaggcatg ggctgcaacc aggggatcgc attactggta ttactgacgg tggaattgtg tttatcatca ttatgattga ggtaaaaatc gcgcgcaatt aaagtatatt tccaagaagt ttagctcact ctctagaact tgacctcgag gtacatttat ttattaatag gtcaataatg gaccttatgg tacataactt ggtggagtat tacgcccct ggtgatgcgg tccaagtctt ctttccaaaa gtgggaggtc gtgatgctgg ctaaaagcaa aaaatcagcg aacttgttaa tcatgctgct ccttgaatac taagtcgagt tcagcaactt cggcaaagga gtcctgccta agaaacaagg acttactcgt cagcaattaa tctgccttag caacagttcg ggaaattatg aaaggtgcct cttatcccta acttaaagct taatgatttt gtatgttgtg gtaatgtgag attacgccaa gtggcggccg tegatacege gcccattgac atcataatat tattgactag agttccgcgt gacgtcaatg atatgtcaag cccaqtacat ctattaccat cacggggatt atcaacggga ggcgtgtacg ggagacgcca gcggccggga cctatagact cgttttgata ctcattgtca tggaaaccta aagaggctga aaaggccgaa gacttgaaaa gttcatgctc tactcagcgt dcacccaaac tggtactggt aacgtactct acaagggaag tacgetttgg gttggtgacc aaatggttgg aagagaacca tatgatcccg ctgttataag taataaaac aaaaaattag atgaatccc cttccggctc tatgaccatg ctccaccgcg atcaagctta tgacattgat ccatatatgg ttgtatccat caagtgtatc tggcattatg tggcaccaaa aacgacccc acttccatt ttagtcatcg cggtttgact atgggcggta cagatcgcct tccagcctcc gtaagtaccg caccacaccd tttaggctat acctaaaatc agcggaaac tgaaattcga aaccttccga cagctcagag gattgaggga cagaaatcga ctgatttaac gaagctgggt atctcgctct tactcagtgc ttattgataa ctacacaata atgaacttga aggaaacagc cacacatggt cgggaaaagc aaattgatcc ctgagcactg catcttgtca acactttatg aaagctggag ggaattcgat ctgaccgccc gccaataggg attgacgcaa caagagtgac attgcatacg gagtttgttt accgccatgt agttcatage ggcagtacat atggcccgcc catctacgta gcgtggatag agtgaaccgt ccgggaccga tattgtataa taccgagtaa aatccgttga actcaaagac accacccgtc agattgaaga acttacctgt gccgaacgag taacatgttg ggcattctgg attggttgga cagacctagg ctaacacagt aaaatttatt teegggatet aaaaaacaat acaacaaaa cttgattata cattgtcago aatttcacac ccgggctgca gtccaacatt ggtggataca cccaggettt aaagggaaca gctattggcc cggggtcatt gacagaatgg atgacggtaa cttggcagta ttccccgtgc ctcgttgacc ccatagtaac ctgcccactt acatcaatgg acgtcaatgg acteegeeee atagaagaca gcgagcattc ccatggtata tgccaaattc actcattgtc aagcgaatgc cacttccagg ctttaaaaat gtacaatatg tcatctagtc ctagcaacta ctacgccata atgetteaac qaagttttgc ctactagete tctagagcga tgcctacatc tgaacattat gatgcctatc cattaggcac attgacgtca aggtaaacgc agcggataac aaccctcact ttacggtaaa ttttggcagt aatgttaatt attaagttaa agtggatcc catcagattg attggctcat taatcaatta acggtaaatg acgtatgttc ggaacgcgga gactttccta caccccattg tgtcgtaaca tatataagca tttgacctcc 2341 2401 2461 2581 2641 2701 2941 2761 2821 2881 3001 3061 3121 3181 3241 3301 3361 3421 3481 3601 3781 3841 3541 3721 3901 3961 4021 1081 4141 1321 3661 1201 4261 4381

2

25

30

35

15

S

gaacgtette cggccgtatt caagggcca aggcyccctg gtgcaaggtc gaaccaggtc gtgggagagc gtggcggtag taagagtcag ttcctttcca ttcgtgtgtg tatttcgacg gtcaacagtt gggaggcttg tttcagaaac cagattcacc gaaagccgag agttatccct ggccctgggc cttccttagc caacgtgaat tgacaaaact cttcctcttc atgcgtggtg cggcgtggag ccgtgtggtc agggcagccc gaccattatt gctctttgcc gatggaagac gcagtactcg atctcgggta gagccctggt gtggaggcca cacggactct aacaacgccg cgtggaactc caggactcta ctgaggtcac ggtacgtgga aggagtacaa gggcctata gctcctaaca gcacaaggcc ggctgacgca tgtattctga tgtagtctga tegtgactte aaatttaata ttctggcttt aggagtcggg ctggattcac tggagtgggt cagcetecae ggggcacagc cctacatctg gaccgtcagt ccaaagccaa tegeegtgga ggcagcaggg cgcagaagag ccataacatg ttccacatcc ccgtgaaagg tgaatagcct catttggggg ccaaatcttg acagcacgta agctgaccaa tgctggactc gtgggttatt cagagactga tcacatatac ctccacgcga taacagactg ggctctgcac aaccactaca gggaagggc tatgctgcac ggcacccaga gagcagtaca tgctatactg tttttggctt gcctataggt ctctgtcctt agcgtgggat gcggcggagc gcagctcctt gggctcgcac gctgagttgt gctgacagac tegttratta tegetttgg gtgcagctgc tgtgcagcct tatctgcaaa attatgataa accgtctcct agcacctctg gtgacggtgt ctacagtcct aaagttgagc ctcctggggg teceggace aagttcaact ctgaatggca aaaaccatct tcccgggatg cccagcgaca acgeeteeeg aagagcaggt cattactaat atttacaaat ccagtgtgcc tggagggag atcattcatc ccctgaggtc aggettetat gacaacagac aaccctggtc ggtggacaag cctcatgatc cccatcgag ctacaagacc caccgtggac cgtcggatca tagagtetee aaacacgtta taccacgggg teeggetgte agcacctgaa cctgccccca cgatactttc tatgccaata cgtggagatt ctgttaacgg agacataata aggtacccag ccaggctcca cagcagcttg gccgcgggag ccaggactgg gtatagctta cccatttatt tattaaacat ttctccggta tggtcgctcg cccaccacca gatgcaggca atttaatatt tttcttcgtg ctcctccaag cccgaaccg caaggatatt atgattcaaa tgcacacctt ccgtgccctc ccaaggacac gcctggtcaa cggagaacaa acatgggctc gctgggtccg ttgatggtgg tatattactg 9999cca999 aggactactt caccgtgccc gccacgaaga ccgtcctgca ccctcccagc aqqtqtacac acagcaagct tgatgcatga aggatggggt cagcacaatg ggcagaagaa cgttgcggtg ctgcagtcac atatatctaa cagagccgaa gggggtccct ccctggcacc gcaacaccaa ccaagacaaa ctattggcta ccgcagtttt cagcggctca gcgcgccacc ctcttatgca ataggtgatg ctattggtga tgaaaatgag teggtettee ccccaaaac agcgtcctca tccaacaag cgagaaccac aatgggcagc ttetteetet tcatgctccg gtatttttac ggtatgtgtc tgggtcttt gtgtttacct atgaatttct teggetgege gcctggatga ccccgaatt tgcctggtca accagcggcg cacacatgcc gtggacgtga gtgcataatg agcctgacct acaactatct cgtgttccgg cccatgcctc gacttaggca ttaaggcagc aggtaactcc gtaaagccgg aaaagcaaaa atctcaagag gacacagccg aacgtggtga cacaagccca tecttatget gaccactccc tececegtge ttgctgccgc 6481 6541 6841 5221 5341 5461 5521 5581 5641 5701 5761 5881 5941 6001 6121 6181 6241 6301 6361 6421 6601 6661 4681 4741 4981 5101 5161 5281 5401 5821 6061 4801 4861 4921 5041

8

25

15

10

30

accctcggtg gccagaaaa gacaatggcc gacagtggcc ctccaaacaa gccctctgcc agtcgttaag ctatgaggac ttactcaact cgtcctaggt gcttcaagcc gtggaagtcc gacagtggcc aggaaacttc tactgccaca ttggctactg cattatccca cgtattacgc ttacccaact aggcccgcac ccaataggcc gagtgttgtt agggrgaaaa ggtcaacgag actcaatggc aaaaaaggcc catatgccct aagcgttaat tgtatccacc aatagatagg atcaagaggg taataataca tcaccaataa ttgtctcctg accaaccttc cctgtaatgc gcagtactgc tactgagtgt actactactg gagatgcatt tggtggtcat ccagctcagg tgacacagcc ccaagctgac cctctgagga cgggagccgt cgcttgagca ccgtggagaa ttgaaggggg CCACCACACC tgtgcagagg gettettge atttttcaag aggcaccaac aaatggatta gcttcaccca tatagtgagt aaccctggcg catttttaa agatagggtt aatagcgaag tggaaattgt ccaacgtcaa agataaataa tgacaagatg ctcttgggtt gatagcagga attggtagtg cagtgctcag ttaaaaactt gcttgccgat aactaagcac ttggtaaggc gacatcactc gggaaaacta ataaatccga aggcttggaa ggccaggate acctgctctg gatgaaggtg tcctatgage caggcccctg ttctctggct agcggagga ggagtggaga ggccagtgcc ttcccacct gactcctacc ctgagcctga gaagggagca tctgaatgtg agggcccttt ggaacactga gtttactgag ttccaaagca ctgtgctttc caattcgccc tgactgggaa cagctggcgt gaatggcgaa ggggattcat gatgatgtgc acctaaaaga tgagcttgaa aaaaaatgcc taaatcagct gaatagaccg aacgtggact cgggatcata gtgacgaaat gaaggtcatc qccatcccaa acceptteca tgtttggata cctgccacaa cagccaataa aattaaaaca actatctgca ccaggtggaa ccctgagaga gccgaagctt gaagtcaggc ggaggtgttc ggtcactctg cgtcaaggcg gagggaggga tectegteac atctaaatgg tctcataagt cagcagctac ggtcacgcat ccccccgcac ccttgctgca gttcaatatt tttgtattat gcccggtacc ccctttcgc aaatttttgt tacaacgtcg tgcgcagcct taaatcaaaa actattaaag cccactactc aaaatcatta agcttatgat catgcgaaaa ggctttttat tctttatcgt acatcatttg gggttaacat ggccaataac ggctattctt tccaaatcca caaatgcact gatgactett tcagtggggc gttatcatag aagcgccaga caggacaac ggtaccagca cctccgggat agtacgcggc ctgcccctc cactggtgtg atagcagccc acagctgcca gttcaccgcg tectetegtg aaatctcaac cctctgccct cccagcacc gagaaatcca tcgagggggg gccgtcgttt gcagcacatc aattcgcgtt acaagagtcc gatttttacc tatttaccgc taacaggtgt tcccaacagt aaatccctta agggcgatgg tatcqcaata atcccaccac ccgtcaggaa agctaaatct tegeggaata tgagcttgag aaaccaataa aacagtaatg gctaaaaagg cccatttagt gcaccttgct tetecgggta gacagcagtg tcagtgtccc tatqtttatt accttgacta tggaaggcag agcaaacgac cagcccaagg aacaaggcca gegecatgae attgctgctt gcaagaaata agcaacaaca cacaaaagct cctgcagaat cacattcctt ctgcccatga agatcctgat ttgcctccc gcgctcactg attttgttaa gaaatcggca ccagtttgga aatttattgc ttttatttga taatcgcctt cgatcgccct accetctatc ttaacttaat tggttatgca tcattatatt tttactccc tccctgtaaa aattaacatt gtaaaacgct aataacagca cgttttttcg aggtaaagcg agatataaac 7021 7081 7141 6961 7201 7261 7321 7381 7441 7501 7561 7741 7801 7981 7621 7681 7861 7921 8041 8101 8161 8221 8281 8341 8401 8461 8521 8701 8581 8641 8821 8761 8881 8941 9001 9061

೫

25

30

35

2

gcgagcggta tattgggcgc cccttcggga ggtcgttcgc agggattttg atgaagttt ctaatgagtg aaacctgtcg cgcaggaaag gttgctggcg aagtcagagg ctccctcgtg cttatccggt agcagccact gaagtggtgg gaagccagtt tggtagcggt cttaatcagt acteceegte aatgataccg cggaaggcc ttgttgccgg cattgctaca ttcccaacga ggcagcactg ggcgtcaata cttcggtcct tgagtactca atttccccga aaaacgttct gtaacccact agaagatcc gaaataataa cctggggtgc gcggtttgcg ttaggatgag caggggataa atcgacgctc gttcggtgta accgctgcgc cgccactggc cagagttctt gegetetget aaggatctca tcagctccgg gaaattgtta tccagtcggg aaaaggccgc ccctggaag ccgcctttct aaaccaccgc tagttgcctg ccagtgctgc agtctattaa acgttgttgc cggttagctc gcgtttctgg cacggaaatg gttattgtct ttccgcgcac ccctttagtg actcacgtta taaattaaaa gttaccaatg accagccagc tcatggttat ctgtgactgg getettgeee tcatcattgg ccagttcgat ccatcatgct aagtgtaaag gcggggagag aggaaccgta caggcgtttc tccggcaaac ccatctggcc agatgcttt cgaccgagtt tcaatgcgct catcacaaa gttcagcccg cgcagaaaa tggaacgaaa tggtctgaca cgttcatcca tcagcaataa gcctccatcc gtcgtttggt atggcttcat gtgttatcac agcttttgtt tttcctgtgt ctgcccgctt cgctcggtcg tccacagaat ggatacctgt aggtatotca cacgacttat ggcggtgcta tttggtatct tagatccttt agtttgcgca tgcaaaaag ttaaaagtgc ctgttgagat agcatcttt actttcacca ataagggcga ttattgaagc atttatcagg caaatagggg gaggtatgta ctgtctattt gtcatagctg cggaagcata gttgcgctca tgactcgctg gcaaaaggcc tagctcttga tccagattta aaaaaaggga gaaaaataaa gaaaagccaa ggattggcta tataccgatc cggccaacgc aatacggtta ataaaqatac geegettace ctcacgctgt cgaaccccc gcagattacg gatcttcacc ggagggctta gccagttaat ccctgacgag cccggtaaga aaggacagta tgacgctcag tgagtaaact aactttatcc ccccatgttg gttggccgca gccatccgta tagcagaact gtgtatgcgg gatettaceg gacccgtaat cattaattgc ggctccgccc gctgtgtgca ttagcagagc aaagagttgg ctacggggtc tatcaaaag tctcagcgat gtggtcctgc taagtagttc ataagtgatg acatacgago attaatgaat cctcgctcac caaaggcggt caaaaggcca cgacaggact tttctcatag ttgagtccaa gctacactag tttgcaagca aaagtatata gctcaccggc tgtcacgctc ttacatgatc tcagaagtaa ttactgtcat tctgagaata ccacacaca aactctcaag actgatcttc aaaatgccgc aatgtattta acaccetect cgtaatcatg ttccgacct ctacgatacg tttttcaata aaaagaccaa taatagcacc gcatcottaa attccacaca agctaactca tgccagctgc tetteegett aacatgtgag ttttccata tggcgaaacc aactatcgtc gtgtagataa atactcttcc gegegettgg tcagctcact cacteteetg agcgtggcgc tccaagctgg ggtaacagga cctaactacg ggttttttg gtcatgagat aaatcaatct gaggcaccta cgagacccac gagcgcagaa gaagctagag accaagtcat cgggataata cgtgcaccca acaggaaggc tacatatttg accttcggaa ttgatcttt ggcatcgtgg tcaaggcgag ccgatcgttg cataattctc tcggggcgaa 10141 9481 9541 9601 9781 9901 10261 10981 9301 9361 9421 9661 9721 9841 1966 10021 10081 10381 10561 10681 11101 10201 10321 10441 10501 10621 10741 10801 11041 11341 11401 11521 10861 10921 11161 11221 11281 20 25 30 35

11581 aaagtgccac

SEQ	Н	NO:101 (pTnMCS (CMV-prepro-HCPro-Lys-CPA)	S (CMV-prep.	ro-HCPro-Ly	s-CPA))		
·	-	1 ctgacgcgc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg	ctgtagcggc ç	gcattaagcg (	3ggcgggtgt c		cgcagcgtga
n	19	1 ccgctacact	: tgccagcgcc	ctagcgcccg	ctcctttcgc	tttcttccct	teetteeteg
	121	1 ccacgttcgc	. cggcatcaga	ttggctattg	gccattgcat	acgttgtatc	
	181	1 tatgtacatt	tatattggct			tgttgacatt	gattattgac
	241	l tagttattaa	tagtaatcaa	ttacggggtc	attagttcat	agcccatata	tggagttccg
Ç	301		cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	cccgcccatt
01	361	l gacgtcaata	atgacgtatg		aacgccaata	gggactttcc	attgacgtca
	421	l atgggtggag		aaactgccca	cttggcagta	catcaagtgt	_
	481	l aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta
	541	catgacctta	tgggacttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac
91	601	catggtgatg	cggttttggc	agtacatcaa	tgggcg+gga	tagcggtttg	actcacgggg
CI	661	-	ctccaccca	ttgacgtcaa	tgggaytttg	ttttggcacc	aaaatcaacg
	721	-	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt
	781	acggtgggag	gtctatataa	gcagageteg	tttagtgaac	cgtcagatcg	cctggagacg
	841	. ccatccacgc	tgttttgacc	tccatagaag	acaccgggac	cgatccagcc	
ć	901	_	attggaacgc	ggattccccg	tgccaagagt	gacgtaagta	ccgcctatag
07	961	. actctatagg		tggctcttat	gcatgctata	ctgtttttgg	cttggggcct
	1021		gcttccttat	gctataggtg	atggtatage	ttagcctata	ggtgtgggtt
	1081	attgaccatt	attgaccact	ccctattgg	tgacgatact	ttccattact	aatccataac
	1141	atggctcttt	gccacaacta	tctctattgg	ctatatgcca	atactctgtc	cttcagagac
96	1201	-	tctgtattt	tacaggatgg	ggtcccattt	attatttaca	aattcacata
<b>C7</b>	1261		ccgtcccccg	tgcccgcagt	ttttattaaa	catagcgtgg	gatctccacg
	1321	cgaatctcgg	gtacgtgttc	cggacatggg	ctcttctccg	gtagcggcgg	agcttccaca
	1381		ggtcccatgc	ctccagcggc	tcatggtcgc	teggeagete	cttgctccta
	1441	acagtggagg	ccagacttag	gcacagcaca	atgcccacca	ccaccagtgt	gccgcacaag
,	1501		tagggtatgt	gtctgaaaat	gagcgtggag	attgggctcg	cacggctgac
20	1561		gacttaaggc	agcggcagaa	gaagatgcag	gcagctgagt	tgttgtattc
	1621	tgataagagt	cagaggtaac	tcccgttgcg	gtgctgttaa	cggtggaggg	cagtgtagtc
	1681	tgagcagtac	tcgttgctgc	ეენენენენე	accagacata	atagctgaca	gactaacaga
	1741	-	ccatgggtct	tttctgcagt	caccgtcgga	ccatgtgcga	actogatatt
ų,	1801		ctctttacca	attctgcccc	gaattacact	taaaacgact	caacagctta
33	1861	acgttggctt	gccacgcatt	acttgactgt	aaaactctca	ctcttaccga	acttggccgt
	1921	aacctgccaa	ccaaagcgag	aacaaaacat	aacatcaaac	gaatcgaccg	attgttaggt
	1981	aatcgtcacc	tccacaaga	gcgactcgct	gtataccgtt	ggcatgctag	ctttatctgt
	2041		cgatgcccat	tgtacttgtt	gactggtctg	atattcgtga	gcaaaacga
	2101	cttatggtat	tgcgagcttc	agtcgcacta	cacggtcgtt	ctgttactct	ttatgagaaa

aggggatcgc tggaattgtg gegegesatt ctttaaaqtq aagaggaaaa gccatgggtt aatcgccctg ttgggacaag cttaggcatg ggctgcaacc tttatcatca ttatgattga attactggta ctctagaact tgacctcgag gtacatttat ttattaataq gtcaataatg ggtggagtat gaccttatgg ggtgatgcgg agccgacctt acatgatate tccaatctca ctcgacacgg tatctattcq cggactaggc aaagtatatt tccaagaagt ttcctgacgg ggtaaaaatc ttagctcact tacataactt tacgcccct tccaagtctt acggtgcatt tattgactag gcccattgac accaatttct aaaatcagcg cggcaaagga caacagttcg ggaaattatg tctgccttag taatgatttt gtaatgtgag gtatgttgtg gtgatgctgg tcagcaactt ctaaaagcaa aacttgttaa gtcctgccta tcatgctgct agaaacaagg acttactcgt aaaggtgcct cadcaattaa cttatcccta ccttgaatac acttaaagct attacgccaa gtggcggccg togataccgc atcataatat agttccgcgt cccagtacat atcaacggga taagtcgagt gacgtcaatg atatgtcaag ctattaccat cacggggatt ggcgtgtacg ggagacgcca gcggccggga tggaaaccta aaaggccgaa gacttgaaaa cgttttgata gttcatgctc gttggigacc aagagaacca cttccggctc tatgaccatg tggcattatg tggcaccaaa cagatcgcct aaagctcatg ctcattqtca tggtactggt aagaggctga tactcagcgt acacccaaac acaagggaag tacgetttgg ctgttataag aaatggttgg taataaaac atgaacttga aaaaaattag atgaatccc tatgatcccg atcaagctta ttgtatccat tgacattgat ccatatatgg actttccatt caagtgtatc atgggcggta tccagcctcc aacgtactct ctccaccgcg aacgacccc ttagtcatcg cggtttgact agcggaaaac ctgatttaac atgttcaaag caccacaccg gaagctgggt tttaggctat atctcgctct acctaaaatc tgaaattcga aaccttccga cagctcagag gcttgcgggc cagaaatcga ctacacaata cgggaaagc tactcagtgc ttattgataa aaattgatcc ctgagcactg catcttgtca acactttatg aggaaacagc aaagctggag ggaattcgat attgcatacg agttcatago ctgaccgccc gccaataggg catctacgta ccgggaccga cacacatggt accgccatgt atggcccgcc attgacgcaa agtgaaccgt gcgtggatag gagtttgttt ggcagtacat actcaaaqac tattgtataa accacccgtc agattgaaga cttgattata attggttgga cattgtcage ctcgttgacc ggtggataca cccaggettt tttcagagca taccgagtaa aatccgttga cagacctagg acttacctgt gccgaacgag taacatgttg ggcattctgg aaaatttatt teegggatet aaaaaacaat acaacaaaa aatttcacac aaagggaaca ccgggctgca gctattggcc gtccaacatt cggggtcatt gcccgcctgg ccatagtaac ctgcccactt atgacggtaa cttggcagta gagctcgttt atagaagaca ctaacacagt acatcaatgg acgtcaatgg acteegeede tgccaaattc agcggataac gcgagcattc gtacaatatg tcatctagtc actcattgtc ctagcaacta aagcgaatgc cacttccagg ctactagete gatgcctatc cattaggcac agtggatccc ccatggtata ctacgccata atgetteaae gaagttttgc tctagagcga ctttaaaaat tgcctacatc aggtaaacgc aatgttaatt attaagttaa tgaacattat aaccctcact catcagattg attggctcat taatcaatta acggtaaatg acgtatgttc ttacggtaaa attgacgtca gactttccta ttttggcagt caccccattg tatataagca tttgacctcc tgtcgtaaca 2341 2401 2461 2521 2641 3181 2581 2701 2881 2941 3001 3061 3121 3301 3361 3421 3481 3541 3601 3661 3721 3901 3961 2281 2821 3781 3841 4021 4081 4141 4201 4261 432I 10 15 35

2

25

ctataggcac caccccccct gaccattatt gatataga cacggactct aacaacgccg atctcgggta gagccctggt gtggaggcca gtggcggtag gatggaagac taagagtcag ttcctttcca ttcgtgtgtg tatttcgacg gggaggettg gcagtactcg gtcaacagtt tttcagaaac cggccgtatt cagattcacc gaaagccgag caagggccca aggcgccctg agttatccct ინნნ.;იიინნ cttccttagc caacgtgaat tgacaaaact cttcctcttc atgcgtggtg cggcgtggag ccgtgtggtc gtgcaaggtc agggcagccc gaaccaggtc gaacgtcttc cctatagact ggggcctata gtgggttatt ccataacatg cagagactga tcacatatac ctccacgcga ttccacatcc gctcctaaca gcacaaggcc ggctgacgca tgtattctga tgtagtctga taacagactg tegtgaette aaatttaata ctggattcac tggagtgggt ttctggcttt ccgtgaaagg aggagtcggg catttggggg cagcetecae cgtggaactc caggactcta ctgaggtcac tgaatagcct gggcacagc cctacatctg ccaaatcttg tegeegtgga gaccgtcagt ggtacgtgga acagcacgta aggagtacaa ccaaagccaa agctgaccaa tgctggactc ggcagcaggg gtaagtaccg atttacaaat tttttggctt cattactaat ctctgtcctt gcggcggagc gcctataggt agcgtgggat gcagctcctt ccagtgtgcc gggctcgcac gctgagttgt tggagggcag gctgacagac atcatccatc tegtttatta ttcgctttgg tatgctgcac tatctgcaaa gtgcagctgc gggaagggc agcacctctg aaagttgagc ctcctggggg tgtgcagcct attatgataa accetctcct gtgacggtgt ggcacccaga tcccqqaccc acagcaagct caccgtggac aagagcaggt ctacagtcct aagttcaact gagcagtaca ctgaatggca aaaaccatct tecegggatg cccagcgaca acgeeteeeg tgctatactg caagagtgac tattaaacat ttctccggta gtatagetta cgatactttc tatgccaata ccatttatt tggtcgctcg cgtggagatt gatgcaggca cgtcggatca tttcttcgtg cccaccacca ctgttaacgg agacataata aggtacccag tagagtetee atttaattt ccaggctcca gacaacagac aaacacgtta taccacgggg aaccctggtc ccctgaggtc ctcctccaag teeggetgte cctcatgatc ctacaagacc ccccgaaccg cagcagettg agcacctgaa gccgcgggag ccaggactgg aggettetat ggtggacaag cccatcgag cctgccccca ttccccgtgc ataggtgatg ctattggtga aggatggggt ctcttatgca ctattggcta ccgcagtttt acatgggctc cagoggotca cagcacaatg tgaaaatgag cgttgcggtg ctgcagtcac atatatctaa caaggatatt cagagccgaa ggcagaagaa gcgcgccacc gctgggtccg gggggtcct ttgatggtgg atgattcaaa tatattactq 9999ccaggg tgcacacctt ccgtgccctc caccgtgccc ccgtcctgca ccetcccage gcctggtcaa ccctggcacc qcaacaccaa ccaaggacac gccacgaaga ccaaqacaaa aggtgtacac cggagaacaa aggactactt gaccactccc gtatttttac ggaacgcgga accetttgg tccttatgct acaactatct teceeegtge cgtgttccgg cccatgcctc gacttaggca ggtatgtgtc ttaaggcagc gtgtttacct aggtaactcc ttgctgccgc teggetgege tgggtcttt gcctggatga atctcaagag atgaatttct gtaaagccgg aaaagcaaaa gacacagccg ccccgaatt teggtettee tgcctggtca aacgtggtga cacacatgcc ccccaaaac gtggacgtga gtgcataatg accagcggcg cacaagccca agcgtcctca tccaacaaq cgagaaccac agcctgacct aatgggcagc ttetteetet 4621 4741 4981 4801 4921 5041 5101 5161 5221 5341 5281 5401 5641 5761 5461 5521 5581 5701 5821 5881 5941 6001 6061 6181 6121 6241 6301 6361 6421 6481 6601 6661 6721

20

25

30

35

gggaggaaac agggccctct agtcgtatta atgtgtatcc cacttgtctc ggataataat gtgaatgatt ggctcaccaa agggacaatg gaccgtccta ggagcttcaa cgtgacagtg accetecaaa gaagacagtg tgctactgcc aagagtcgtt aacttggcta ccacattatc gcgttaccca tgtaagcgtt gttgagtgtt cagggtcaac cttactcaat taaaatagat ctccctgtaa gccaccctcg attgccagaa catctatgag ctgttactca gcagtggaag ttacatatgc aagaggccg taaccastad caaagggcga gataaaaag gttatcaaga cgagcagtac gagatttttc gccaggcacc agagettgee gcaataaatc agaccaccac ttcgcttcac gctcattttt gaaagataaa atagacatca gtgactacta acccgggagc tgacgcttga gcaaaatgga gaatggaaat ccgagatagg atatgacaag tgcttaaaaa gccctcttgg aataactaag caaattggta acagggaaaa cgcagaagag ctggagatgc ctgtggtggt gctccagctc ggaccaagct cctcctctga gcaccgtgga ttttgaagg tgatgtgcag gtggcttctt ccctatagtg gaaaaccctg cgtaatagcg actccaacgt catcagtgct atcgatagca gcattggtaa agctgacaca ggccaggccc agtgactcct cacggarcac tgggtttcct ttggtgacga ccagccatcc ctttcctatg gcgggagtgg tacctgagcc catgaaggga ggaagggccc gcaggccagt cctgaatggc tgttaaatca aaagaataga ctccgggatc ttaggggatt gatgatgatg aaaacctaaa tattgagctt catgaaggtc aacaccggtt ctttgtttgg taaaattaaa actactatct agattctctg cactcigaat aagaacgtgg cgtaaaaat aaccactaca atcacctgct gaagatgaag ttcagcggag ctgttcccac attttccaaa tatctgtgct acccaattcg tegtgactgg cgccagctgg gaggggttaa aacgctaaaa aggcaaatgc ggctctgcac gccgaaaag gatccctgag tagggaggtg gtgtctcata ggccagcagc ccagttcaat tgttttgtat ggggcccggt ttttacaacg agttgcgcag tccactatta tggcccacta accaaaatca gaaagcttat atacatgcga cgcggcttt tcttctttat ataacatcat taatccaaat atgggccaat gctgatgact caccaqccaa gcagaagtca ggcccaggtg ccccgtcaag ccaggtcacg cctccctcgt aacatctaaa accettget gttaaatttt ttataaatca aacggccagg ctcggtcact gcggaggag gtgcccccg atccccttt atttcgcgga gcgatcccac taaaattcgc aatgatttt gctaaaccaa gcaaacagta aaagcacctt gtggttatca aggetgeece ccacactggt cagatagcag gctacagctg aatgttcacc gactcctctc cttcctctgc ataaaatctc cttccccagc tgagagaaat gattaacagg cttgcagcac ccttcccaac gcaaaatccc ggaacaagag atcagggcga atteegteag gcatatcgca tgctatttac tgaagctaaa ccctgagett gacceteegg ctatcagtgg acaagtacgc ccctcgaggg ctggccgtcg tgatgcatga aagcgccaga ccccaggaca attggtacca gagaattaac gacagcaaac tcccacaaaa gcccctgcag gccattgctg aaqcacattc ccattgcctc acttaatcgc caccgatcgc accttaactt ggctggttat gccaatttat aggttttatt gggtcattat ctatttactc ataaataaca tgcaggtaaa ttcagatata aaatatgttt gccaccttga actgacagca ggtcagccca gccaacaagg gcctggaagg caaagcaaca ttcgcgccat acagcaagaa ctgctgccca cctagatect cgcgcgctca aatattttgt gccgaaatcg aaaaccgtct acagtaaaac taatccctgt tctccgggta gtgtcagtgt gttccagttt 7561 7141 7201 7501 7621 7681 7741 7801 7861 7921 7981 8041 8101 8221 8281 8341 8401 8581 8641 8761 7021 7081 7261 7321 7381 7441 8161 8461 8521 8701 8821 8881

2

30

23

35

15

atcacgattt ttatccgctc aagctccctc taatcaacaa tgcctaatga gggaaacctg gcgtattggg gcggcgagcg taacgcagga cgcgttgctg ctcaagtcag tctcccttcq gtaggtcgtt cgccttatcc ggcagcagcc cttgaagtgg gctgaagcca tcaagaagat cgctggtagc ttaagggatt aaaatgaagt atgcttaatc ctgacicccc tgcaatgata agccggaagg taattgttgc cggttcccaa tatggcagca tggtgagtac tgccattgct ctccttcggt tgggtgagca tctcatgagc cccggcgtca tggaaaacgt atgttgaata gatgtaaccc gctattcatc ctttccagtc caaaggcttg gctgaaataa gttcccttta tgtgaaattg gaggcggttt tegttegget aatcagggga gtaaaaaggc aagcctgggg aaaatcgacg ttececetgg tgtccgcctt ccgaccgctg tcagttcggt tategeeact ctacagagtt tctgcgctct aacaaaccac aaaaaggatc aaaactcacg ttttaaatta acagttacca ccatagttgc gccccagtgc taaaccagcc tccagtctat gcaacgttgt aagcggttag tttctgtgac cattcagctc gttgctcttg cactcatggt tgctcatcat gatccagttc cgacacggaa agggttattg ccagcgtttc tcgcccattt agtggctatt cttcctgcca caagacccgt aatgaaaagc caaccatcat ctgtttcctg ataaagtgta tcactgcccg cgcgcgggga ttatccacag gagcatcaca tgtaggtatc cccgttcagc ctatcaatgc atcagctttt ctgcgctcgg gccaggaacc taccaggcgt accgga tacc gtaggcggtg gtatttggta tgatccggca ttgtgcaaaa agacacgact acgcgcagaa cagtggaacg aaggatotto acctagatoo gatotgtota tttogttoat ttaccatctg acttggtctg ttatcagcaa tgcaacttta tccgcctcca aatagtttgc ggtatggctt gcagtgttat gtaagatgct ataccgcgcc acatagcaga actttaaaag cggcgaccga ccgctgttga ggaataaggg agcatttatc tttactttca tgagttgaga atggtcatag agccggaagc aatcggccaa actataaaga cctgccgctt caacccggta ggattagcag agcgaggtat gctggattgg atgtataccg cactgactcg ggtaatacgg ccagcaaaag ccccctgac tageteaege gcacgaaccc gtctgacgct tagaaggaca gaaaaagagt tggtagctct gcagcagatt tctaaagtat atatgagtaa cacgeteace ggetecagat acgggagga ttcgccagtt ctcgtcgttt atcccccatg taagttggcc catgccatcc ccaactgatc ttcagcatct ggcaaaatgc cgcaaaaag atattattga atagtgtatg gaaaactctc aaggatctta taaataagtg acaacatacg tgcattaatg tggcgtaatc tcacattaat cttcctcgct gtcttgagtc accacaccgt actcaaaggc gagcaaaagg ataggeteeg aggtggcgaa acccgacagg ctgttccgac cgctttctca tgggctgtgt acggctacac ttgtttgcaa tttctacggg gattatcaaa ctatctcagc taactacgat gaagtggtcc gagtaagtag tggtgtcacg gagttacatg ctcttactgt cattctgaga ttgtcagaag tccttttca tgtaaaagac acaattccac tgccqttttt attgcgcgct tegtgeeage ctgtaatagc atggcatcgt gtgagctaac cgctcttccg gtgcgctctc tggcctaact gtatcagctc gcgtttttcc ggaagcgtgg cgctccaage aagaacatgt ggtaactatc actggtaaca cctttgatct gttaccttcg ggtggtttt ttggtcatga tttaaatcaa agtgaggcac gtcgtgtaga ccgcgagacc gccgagcgca cgggaagcta cgatcaaggc actegtgeae acaggcatcg cctccgatcg atacgggata tcttcggggc ctgcataatt tcaaccaagt aaaacaggaa ctcatactct 9481 9301 9361 9421 9541 9601 9661 9721 9781 9841 9901 1966 10201 10021 L0081 10321 10441 10141 10261 10381 10501 10561 10681 10741 10621 10801 10981 11041 10861 10921 11101 11161 11461 11221 11281 11401

25

30

35

2

S

11521 ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc					*	2 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	
	cg cacatttccc	gggttccg	aaacaaatag	ttagaaaaat	ttgaatgtat	ggatacatat	11521

	cgcagcgtga tcctttctcg catatcataa gattattgac	tggagttccg cccgcccatt attgacgtca atcatatgcc	tcgctattac actcacgggg aaaatcaacg gtaggcgtgt cctggagacg	ccgcggccg cctatag cttggggcct ggtgtgggtt aatccataac cttcagagac	gatetecaca agettecaca ettgetecta geogeacaag eaeggetgae tgetgtatte	cayuyaayo gactaacaga acttgatatt caacagctta acttggccgt attgttaggt cttatctgt gcaaaaacga
	ggtggttacg tttcttcct acgttgtatc tgttgacatt	agcccatata cccaacgacc gggactttcc catcaagtgt gcctggcatt	gtattagtca tagcggtttg ttttggcacc caaatgggcg cgtcagatcg	cgatccagcc gacgtaagta ctgtttttgg ttagcctata ttccattact atacttgtc	attatttaca catagcgtgg gtagcggcgg tcggcagctc ccaccagtgt attgggctcg gcagctgagt	atagotgaca ccatgotgaca ccatgotgaca taaaacgact ctcttaccga gaatcgaccg ggcatgotag atattcgtga
aaacaaacag	-CPA)) cggcgggtgt ctctttcgc gccattgcat attaccgcca	attagttcat tggctgaccg aacgccaata cttggcagta taaatggccc	gtacatrtac tgggcytgga tgggagtttg cccattgacg tttagtgaac	acaccgggac tgccaagagt gcatgctata atggtatagc tgacgatact ctatatgcca	ggroccarre tttattaaa ctcttctccg tcatggtcgc atgcccacca gagcgtggag	grycrycrad accagacata caccgtcgga gaattacact aaaactctca aacatcaaac gtataccgtt gactggtctg
ttgaatgtat ttagaaaat cac	(CHOvep-prepro-HCPro-CPA):tgtagcggc gcattaagcg cggc; tgccagcgcc ctagcgcccg ctcc cggcatcaga ttggctattg gcca	ttacggggtc atggcccgcc ttcccatagt aaactgccca	ctacttggca agtacatcaa ttgacgtcaa acaactccgc gcagagctcg	tccatagaag ggattccccg tggctcttat gctataggtg cccctattgg	tacaggatgg tgcccgcagt cggacatggg ctccagcggc gcacagcaca	cocgregory cgcgcgcgcc tttctgcagt atctgcccc acttgactgt aacaaaacat gcgactcgct tgtacttgtt
cac	(CHOvep-pr ctgtagcggc tgccagcgcc cggcatcaga tatattqqct	tagtaatcaa cttacggtaa atgacgtatg tatttacggt cctattgacg	tgggactttc cggttttggc ctccacccca aaatgtcgta gtctatataa	tgttttgacc attggaacgc cacacccctt gcttccttat attgaccact gccacacacta	tctgtatttt ccgtccccg gtacgtgttc ggtcccatgc ccagacttag tagggtatgt gacttaaggc	cagaggicaac tcgttgctgc ccatgggtct ctctttacca gccacgcatt ccaaagcgag tccacaaaga cgatgcccat
ggaracarar cgaaaagtgc	ID NO:102 (pTnMod 1 ctgacgcgcc c 61 ccgctacact t 121 ccacgttcgc c	tagttattaa cgttacataa gacgtcaata atgggtggag aagtacgccc	catgacctta catggtgatg atttccaagt ggactttcca acggtgggag	ccatccacgc ggaacggtgc actctatagg atacaccccc attgaccatt atggctcttt	tgacacggac tacaacaacg cgaatctcgg tccgagcct acagtggagg gccgtggcgg	tgataagagr tgagcagtac ctgttccttt ttacatgatt acgttggctt aacctgccaa aatcgtcacc tcgggcaata
11581	SEQ ID NO. 1 61 121 181	241 301 301 421 481	541 601 661 721 781	841 901 961 1021 1081	1201 1261 1321 1321 1381 1441 1501 1501	1621 1681 1741 1801 1861 1921 1981
	'n	10	15	50 50	30 73	35

ttatgagaaa ctttaaagtg acatgatatg aagaggaaaa tccaatctca ctcgacacgg ttgggacaag cttaggcatg agccgacctt gccatgggtt tatctattcg cggactaggc aatcgccctg ggctgcaacc ataatgatcc gtttttgtg gtgccttcct agcaaggggg ggtacctctc gttggtctgc attgcatcgc tetetetete tgaccaaagg ataagcagca aaaaccttat attagccttg aaccaactta tecectaatg teceggtaat ggctcgtatg ccatgattac ccgcggtggc gcttatcgat tccaaaggag tccattcgac tettteette caagttctga tcatgctgct tcagcaactt ctaaaagcaa aaaatcagcg aacttgttaa gtcctgccta accaatttct gtgatgctgg taagtcgagt cggcaaagga aaatgaggaa agaaacaagg caacagttcg acttactcgt ggaaattatg ctgtgtgttg ccctcccc ggggcagcac gggctctatg cetetetete attgacccgg agtgcctgtt ttaacaaatg gataataata cttgaaaaaa gatccaagag tgtcatatga cactgatgaa caaacagaga ttatgcttcc acagctatga tggagctcca tgaactcaca gcaatggcaa aatggttgct tcgatatcaa aacaaaatct tggaaaccta cacggtcgtt aaagctcatg tggtactggt aagaggctga aaaggccgaa tactcagcgt acacccaaac cgttttgata gttcatgctc асааддуаад ctcattqtca gacttgaaaa aacgtactct tacgetttgg ctctagagat ctgttgtttg tttcctaata gggatgcggt tctctcggta gtgctgaaga acaattactc attgatgcct acatcacaac aaaaactgat ttatattatt cacacaggaa 9999t9999t ttggaatgaa tcagcaaatt tgaccctgag gctttacact gaaaaatgcc aggtggacta atacacatct gaacaaagc ctgcaggaat tetetecatg gcgacaaaga caaagaggag ggggaaacaa accaaggtat tgcgagette agtegeaeta tttaggctat gaagctgggt agcggaaaac tattgtataa atctcgctct atgttcaaag caccacaccg acctaaaatc tgaaattcga aaccttccga cagctcagag gcttgcgggc ctacacaata cagaaatcga cacacatggt tgccagccat tetetetete teggtaceag attatcttga gttaaggtgg aagatcagag cccactgtcc tctattctgg aggcatgctg aaaataaaa gaagtgatge ctatcattgg aacgccattg taattctcgt ggcacccag tcactaaagg ataacaattt gaactagtgg atcccccggg tggctacagc gacctgagca aaatgatttc tcttcaaaca ctcacaatat taatgtcaaa attgtagtgg tttcagagca aatccgttga cagacctagg actcaaagac taggtgtcat agacaatagc tototototo tctctctctc taccgagtaa accacccgtc agattgaaga gccgaacgag aaaatttatt tggctaataa aggtgccact acttacctgt taacatgttg ctaacacagt ggcattctgg gccttctagt catcacttta atatttgaac tggtaaggta gacggaatgt aaatcattaa ttgtgagcgg aatttccaca tcactcatta ccatcactga tegagetgea caattaaccc tacctqattt cttatggtat gegtteeege ccatggtata tcatctagtc gcgagcattc tgccaaattc cacttccagg gaagttttgc gatctgctgt tgaccctgga attgtctgag gtacaatatg actcattgtc ctagcaacta aagcgaatgc atgetteaae ctactagete agatcacttc tctctctctc ctacgccata aggattggga tetetetete tgccttttat attaattatg aatacattac attttggtaa ttgtgtggaa gccaagcgcg cttagaagt ccctatccaa aagctttcct gtgagttagc atgaaaaggc ggccgctcta cttgacctga tgctccttcc accgctgacc cagagagaa attcatctgt 2281 2341 2401 2461 2521 2581 2641 2701 2761 3181 2821 2881 2941 3001 3061 3301 3121 3241 3361 3421 3481 3541 3601 3661 3721 3841 3961 3781 3901 4021 4081 4141 **4201** 1261 1381

25

30

35

15

tgcactacca gctagtggta atgagactgt agagccttag ttgtactgct gcatttcata caagattttc atgaaacaaa ataactgaaa ccagctgaat ctttgaggaa ttgtatggag ctctagtctc ccttccaact atatatottt ctgaagaatt tcagaatggt cccataaaaa aaataatagg tgtagataca ataagcacac tctgttttct tatttcccag cacattgtta attgaaacta ttctgaaggg agtggatcaa ttcttcgtgt agagtctcct caggctccag aacacgttat accctggtca tcctccaaga cccgaaccgg tttaatattt ggtacccagg acaacagact gtctttattc aaagggcagc acaaaactat tcccaggtaa agactgacat taaaaacaaa aaatattatt tttccacagt tatgtatgtc catggaagat ttggtttgca agcattcaca ctgctcttca tttgtcaaga aaccatagta cccatattt acatgagcaa aaatggacct aaggtcaaac gggctcagcc tatatctaaa aaggatattt ggggtccctt ctgggtccgc gggccaggga ggactacttc aaaaagcagg ctaaacttca tttacagagg aatagcttct agatgcacgt aggcagaaaa agagccgaaa tgatggtggg tqattcaaaa atattactgt gaacactcct aagcttctgc ttgccaccta aagacgtaat cagtgatctc aaattgagta tttgttataa gcctggtcaa ctggcatagg acaatgtgtg tccagcaaa agaatgtaga caaacagtgc aatacajaac ttaagttatt tttctcctat tgttacctac ctctaaggtt gctcaatgga attcaatcca tatggtgtca tgtttaccta tgaatttctc aggatgagaa aggteteega ctatcatcac atggcactga ctcgaaccat gaaagtgcgt agcatggctc ttagcagaac atattcccca taaagccggg cctggatgag aaagcaaaat tctcaagaga acacagccgt cccgaattg cggtcttccc aatgaaatgc gggcgatcca tctctctaga gcattgtcaa ggaagcacat atagaagcaa ctattatttc atgattgtcc tgccaggcca aagaggcttt tataactttc ggtgtagaca cttagcaagg acaaaagaaa atttttaaat aacacatttq tgcttcttta tgtcctgcct acttctggtc tgttttccat tttgctctcc ggccgtatta gttatccctc acagagattg agaaaaaag ctataggaaa tgtaactgaa tcctgatgga aaatgatttc tegtgtgtgg agattcacca aaagccgagg gccctgggct ttcaggctat atttcgacga tcaacagttt ggaggcttgg ttcagaaacg aagggcccat gtgattggat aaagtgcttt acagacccac gtgctgggca tgttgcttac cgtgacttct cttgggacaa agtggctgaa agteeteaga aaggagagaa tttgtcaatt ctctgtcatc tgggtcacaa tggattcact ggagtgggtc atttggggga ageetecace acatttactg attctgctta ttttcccc attgtatatt tcatatcata ctgtatcctc aatgaaacag ttgctaatta ggagtcgggg cgtgaaaggc gaatagcctg gggcacagcg tgatacctgt aatattccag ttcacatgca ttctcactta attatctggt catccaacag aaaactaata caatcccatt aatttaatat tetggetttg qtaqqtttta ttttggataa tggtttaggg ttgcacagct gcaagaagat ggtagagata taggaaagta atqttqtact ttctttactg tatatttgct ctatgttttg gatagattta agaaggatca cgtttattaa ttatqataac acgtatacta ctgacctttt agactttctt aaaaatcagt tcttgtctta tcacaaaagg ttcacaattc cactgcaagt agttggctcc tcacgtcttg taactaataa taaagatccc tettetetee cccagaatta aaatctaacc aacctgtggg tegetttggt gtgcagcctc ggaaggggct atctqcaaat tcattcatct tgcagctgca atgetgeace gcacctctgg 4741 4681 4801 4861 4921 4981 5041 5101 5161 5221 5281 5341 5401 5461 5521 5581 5641 5701 5761 5821 5881 5941 6001 1909 6121 6181 6241 6301 6481 6541 6601 6361 6421 6661

2

25

30

15

egtgeeetee ageagettgg gtggacaaga ccqqctqtcc gcacctgaac ctcatgatct cctgaggtca ccgcgggagg caggactggc cccatcgaga ctgcccccat ggcttctatc tacaagacca accgtggaca gctctgcaca gccaggatca ccgaagcttt aagtcaggcc cctgagagat caggtggaag gaggtgttca ctcataagtg gtcactctgt agcagctacc gtcacycatg agggaggaa ccccgcacg gtcaaggcgg actagtaact tctaaatggg cttgctgcag ttgtattatc ttcaatatt cccggtaccc ceettegee gcgcagcctg acaacgtcgt aatttttgtt gcacaccttt caacaccaag accgtgccca caaggacacc ccacgaagac caagacaaag cgtcctgcac ceteceagee ggtgtacacc cctggtcaaa ggagaacaac cagcaagete agcgccagag gatgcatgag ctccgggatc cagtggggcc aggacaaacg gtaccagcag ttatcatagg tgccccctcg actggtgtgt tagcagccc gtacgcggcc ttcaccgcgg cctctcgtgc aatctcaaca cagctgccag ctctgccctt cccagcaccc tgcctccct cgagggggg ccgtcgtttt ttttgttaaa attcgcgtta cagtttggaa caagagtcca agaaatccag aacaggtgtt cagcacatcc cccaacagtt aatcccttat acgtggtgac acaaqcccaq cccaaaacc tgcataatgc gcgtcctcac ccagcggcgt tggacgtgag ccaacaagc atgggcagcc acacatgccc gagaaccaca gcctgacctg tcttcctcta ctccgygtaa cagtgtcccc ccttgactat acagcagtgg gtcctaggtc agcccaaggc ggaaggcaga ttgctgcttc tgcccatgag catgctrcgt atgtttattg gcaacgacc acaaggccac gcaacaacaa ctgcagaatg caagaaataa gatcctgatt acaaaagcta acattccttc cgccatgact cgctcactgg aatcgccttq gategeeett aaatcggcaa ttccttagca aacgtgaatc cgtgtggtca ggcgccctga gacaaaactc ttcctcttcc tgcgtggtgg ggcgtggagg tgcaaggtct tgggagagca ccctcggtgt gggcagccc aaccaggtca gacggctcct ccagaaaat tactcaactg aacqtcttct ctctccctgt tatgaggaca acaatggcca cttcaagcca acagtggcct gtcgttaagc tccaaacaaa tggaagtccc acagtggccc ccctctgcca attatcccat ggaaacttcg actgccacag tggctactgc tacccaactt agcgttaata atatgcccta gtattacgcg caataggccg agtgttgttc ggcccgcacc ctacatctgc cagcacgtac gtggaactca aggactctac accgtcagtc tgaggtcaca gtacgtggac caaatcttgt ggagtacaag gctggactcc gcagcagggg ggtggtcatc caaagccaaa gctgaccaag cgccgtggag gcagaagagc gacacagcca agatgcattg ctactactgt cageteaggg caagctgacc ctctgaggag gggagccgtg gcttgagcag cgtggagaag tgaaggggga gtgcagaggg cttctttgct ttttcaaga aatggattac caccacaccc cttcacccac atagtgagtc atagcgaaga ggaaattgta gatagggttg ggcaccaact accctggcgt attttttaac aagttgagcc tacagteete tgacggtgtc gcacccagac tcctgggggg cccggaccc agttcaactg tgaatggcaa aaaccatctc agagcaggtg agcagtacaa cccgggatga cctatgagct ccagcgacat cgcctcccgt cctgctctgg aggcccctgt tetetggete atgaaggtga gagtggagac accactacac teccacete gcggaggac actcctaccc tgagcctgac aagggagcac ctgaatgtgg tttcctgaga tccaaagcaa agctggcgta gggccctttt gaacactgat gccagtgcca aatggcgaat aatagaccga tgtgctttcg aattcgccct gactgggaaa aaatcagctc 7081 7141 7201 7021 7261 7321 7381 7441 7501 7561 7621 7681 7741 7801 7861 7921 7981 8041 8101 8281 8341 8461 8161 8221 8401 8521 8581 8641 8701 8761 8821 8881

25

30

ttgcgctcac gctttttatt ggttaacatg atgactctt gctattcttc gattggctat ggaagcataa gactcgctgc atgcgaaaaa ctttatcgta catcatttgg ccaaatccag gccaataaca agccaataaa aaatgcacta aaaagccaac ataccgatca ggccaacgcg atacggttat cctgacgagc taaagatacc tcacqctqta ccggtaagac aggtatgtag agctcttgat cagattacgc gagtaaactt aaatcattag gcttatgatg tcatagctgt caaaadcca ccgcttaccg gaacccccg aggacagtat atcttcacct tgtctatttc gagggcttac gacgeteagt gggcgatggc atttaccgcg cgcggaataa taagtgatgt atttttacca catcaddaaa atcgcaatac gctaaatctt gagcttgagg aaccaataat acagtaatgg caccttgctg teccaece ctaaaaaggc ccatttagtg accegtaatg caccgtgctg gtaatcatgg catacgagco attaattgcg ttaatgaatc cageteacte aaaggeggta acatgtgagc aaaaggccag gataagaaaa gacaggacta teegaceetg ttctcatage ctgtgtgcac tgagtccaac tagcagagcg ccttcggaaa aagagttggt tacggggtct ctcagcgatc tacgatacgg ctcgctcact ctacactaga ttgcaagcag atcaaaaagg aagtatatat tttatttgaa cttccgcttc atttattgct gctaactcac gccagctgca gtaacaggat aatcaatcta ggttatgcat cattatattt taaaacgcta ataacagcaa ccctgtaaag ggtaaagcga catcgttaaa gcgtggcgct tgtagataac ccgtctatca attaacattc ttactcccct aaagaccaag ttccacacaa ctaactacgg tgatcttttc aggcacctat taacttaatg gatataaacg gttttttgc aatagcacca cgcgcttggc ttttccatag ggcgaaaccc gctctcctgt ccaagctggg actatogtot gttttttgt tcatgagatt gggcgaaaaa aaaaaggcca tcaagagggt tgtctcctgt atgattataa cagtactgcc actgagtgta cgatttctgt gtcaacgaga ctcaatggct atagataggt aataatacag ccaaccttca caacaaatgg gcaggaaaga ttatccggta gcagccactg gggattttgg gtatccacct caccaataat ctgtaatgca gggttaattg ccgctcacaa taatgagtga aacctgtcgt attgggcgct cgagcggtat ttgctggcgt ccttcgggaa aagccagtta ggtagcggtg tgaagtttta ttaatcagtg ctccccgtcg agtcagaggt tecetegtge gtcgttcgct aagtggtggc gaagatcctt caacgtcaaa cttgccgata gataaataaa tcttgggtta actaagcact atagcaggat ttggtagtga tggtaaggct acatcactcc ggaaaactaa taaatccgag ggcttggaat ctggggtgcc ccagtcggga cggtttgcgt aggatctcaa ttaccaatgc taaaaactta aaataataat cctttagtga aaattgttat teggetgegg aggggataac aaaggccgcg cgcctttctc ttcggtgtag gccactggca agagttcttg ctcacgttaa agttgcctga gacaagatgt agtgctcagg ttcatcatca tegaegetea ccctggaagc ccgctgcgcc cgctctgctg aaccaccact aaattaaaaa acgtggactc cctaaaagag gagcttgaaa tgacgaaata gtttggatag ggaacgaaaa ggtctgacag gggatcatat gggattcatc atgatgtgct aaaaatgccc aaggtcatcg ccatcccaaa ccggttgcat attaaaacag ctatctgcaa ctgccacaa catcatqcta caatgcgctg gettttgtte ttcctgtgtg agtgtaaagc tgcccgcttt cggggagagg ccacagaatc ggaaccgtaa atcacaaaaa aggcgtttcc gatacctgtc ggtatctcag ttcadcccga geggtgetae ttggtatctg ccggcaaaca gcagaaaaa agateetttt gttcatccat gctcggtcgt acqacttatc 9541 9841 9901 10261 10321 10861 9241 9301 9361 9421 9481 9661 9721 9781 9961 10081 10201 10381 10441 10501 10561 10621 10681 10741 10921 10981 11041 11281 10141 11101 11221 11401 10021 10801 11161 11341

2

25

30

35

15

ccagatttat actttatccg ccagttaata tcgtttggta cccatgttgt ttggccgcag ccatccgtaa tgtatgcggc agcagaactt atcttaccgc gcatctttta	tattgaagca aaaaataaac	cgtga tctcg cataa ttgac	ocatt Ogtca atgcc vagta	19999 19999 19acg 19acg	gecg gect gect ggtt taac taac agac cata
	tattg aaaaa	cgcagcgtga tcctttctcg catatcataa gattattgac	cccgcccatt attgacstca atcatatgcc atgcccagta	actcacggg aaaatcaacg gtaggcgtgt cctggagacg	recgeggeeg cegectatag cttggggeet ggtgtgggtt aatcataac cttcagagac aattcacata
ctcaccggct tggtcctgca aagtagttcg gtcacgctcg tacatgatcc cagaagtaag tactgtcatg ctgagaatag cgcgccacat actctcaagg	ttttcaatat atgtatttag	ggtggttacg tttcttcct acgttgtatc tgttgacatt	ccaacgacc gggactttcc catcaagtgt gcctggcatt	tageogetttg ttttggcacc caaatgggcg cgtcagatcg	ugacicages gacgtaagta ctgittitigg ttagectata ttecattact atactetgic attattiaca catagegigg
			CCCas GGGaC CatCS GCCTC	taged tttg caaat cgtca	cyarc gacgt ctgtt ttagc ttcca atact attat
gagacccacg agcgcagaag aagctagagt gcatcgtggt caaggcgagt cgatcgttgt ataattctct ccaagtcatt gggataatac cggggcgaaa gtgcacccaa	tactcttcct acatatitga aagtgccac	LYS-CPA)) cggcgggtgt ctcctttcgc gccattgcat attaccgcca	tggctgaccg aacgccaata cttggcagta taaatggccc	tgggcgtgga tgggagttgg cccattgacg tttagtgaac	acaccygygac tgccaagagt gcatgctata atggtatagc tgacgatact ctatatgcca ggtcccattt
		-LYS- cggc ctcc gcca atta	•	tgggc tgggg cccat tttag	tgcca gcatg atggt tgacg ctata ggtcc
atgataccgc ggaagggccg tgttgccggg attgctacag tcccaacgat tccgacctc gagtactcaa gcgtcaatac aaacgttctt taacccactc	tgaatactca atgagcggat tttccccgaa	epro-HCPro gcattaagcg ctagcgcccg ttggctattg catgtccaac	atggccgcc ttcccatagt aaactgcca tcaatgacgg	agtacatcaa ttgacgtcaa acaactccgc gcagagctcg	occatagada ggattcccc ggattcccc gcatataggt cccctattgg tacagatgg tacaggatgg tgcccgcagt
		repro gcat ctag ttgg	atgg ttcc aaacl tcaat	agtac ttgac acaac gcage	ggatt ggatt ggtat cccct tctct tacag
cagtgetgea ccagceage gtctattaat cgttgttgec cagctcoggt ggttagetec catggttatg tgtgactggt ctttgeceg catcattgga cagttcgatg	acggaaatgt ttattgtctc tccgcgcaca	ID NO:103 (pTnMod (CHOvep-prepro-HCPro-LYS-CPA))  1 ctgacgcgc ctgtagcggc gcattaagcg cggcggtg 61 ccgctacact tgccagcgcc ctagcgcccg ctccttccg 121 ccacgttcgc cggcatcaga ttggctattg gccattgca 181 tatgtacatt tatattggct catgtccaac attaccgcc 241 tagttattaa tagtaatcaa ttaccgacc attacttca	cttacggtaa atgacgtatg tatttacggt cctattgacg	cggttttggc ctccacccca aaatgtcgta gtctatataa	attggaacge actggaacge gctcccttat attgaccact gccacacta tctgtatttt ccgtcccccg
		d (CH ctgt tgcc cggc tatai	cttac atgac tatti cctat	cggtt ctccs aaatc gtcta	attgg cacac gcttc attga gccac ccgtc
catctggccc cagcaataaa cctccatcca gtttgcgcaa tggcttcatt gcaaaaaagc tgttatcact gatgcttttc gacgagttg taaaagtgct tgttgagatc	taagggcgac tttatcaggg aaataggggt	103 (pTnMoc ctgacgcgcc ccgctacact ccacgttcgc tatgtacatt	cgttacataa gacgtcaata atgggtggag aagtacgcc catgacctta	catggtgatg atttccaagt ggactttcca acggtgggag ccatccacac	ggaacggtgc actctatagg atacaccccc attgaccatt atggctcttt tgacacggac
=		:103 ctgac ccgcl ccac tatgt	cgtte gacgt atggg aagta catga	catgg atttc ggact acggt	gggaac actct ataca attga atggc tgaca
11461 11521 11581 11641 11701 11821 11881 11981 11941 12001	12181 12241 12301		301 361 421 481 541	601 661 721 781 841	901 961 1021 1081 1141 1201 1261
		Ōas			
5 10	15	20	25	30	35

tgttgtattc ctttaaagtg aagaggaaa acatgatatg gtttttgtg cttgctccta gccgcacaag cacggctgac cagtgtagtc gactaacaga acttgatatt caacagctta attgttaggt gcaaaaacga ttatgagaaa agccgacctt tccaatctca ctcgacacgg gccatgggtt tatctattcg cggactaggo aatogoootg ttgggacaag cttaggcatg ataatgatco gtgccttcct attgcatcgc agcaaggggg ggtacctctc tototototo ataaqcaqca aaaaccttat attagccttg acttggccgt ctttatctgt ggctgcaacc tgaccaaagc ccatgtgtga gaatcgaccg ctgttactct ctaaaagcaa acacccaaac aacttgttaa gtcctgccta ggaaattatg ggggcagcac gggctctatg teggeagete ccaccagtgt cggtggaggg taaaacgact ctcttaccga ggcatgctag atattcgtga accaatttct gtgatgctgg taagtcgagt tcagcaactt aaaatcagcg cggcaaagga tcatgctgct ctgtgtgttg aaatgaggaa cetetetete agtgcctgtt ttaacaaatg qataataata cttgaaaaa attgggctcg gcagctgagt atagctgaca agaaacaagg caacagttcg acctacaca attgacccgg acttactcgt gtgctgttaa cacggtrgtt tggtactggt aagaggctga gacttgaaaa gttcatgctc ttggaatgaa tcatggtcgc atgcccacca gagcgtggag gaagatgcag accagacata caccgtcgga gaattacact aaactctca aacatcaaac gtataccgtt gactggtctg aaagcccatg ctcattgtca tggaaaccta aaaggccgaa tactcagcgt cgttttgata aacgtactct acaagggaag tacgetttgg ctctagagat ctgttgtttg tttcctaata 9999t9999t gggatgcggt tctctcggta gtgctgaaga acaattactc aaaaactgat ttatattatt ctatcattgg cggacatggg gtctgaaaat tcccgttgcg cacacacacc tttctgcagt gcgactcgct atgttcaaag gaagctgggt agcggaaaac tttaggctat acctaaaatc tgaaattcga aaccttccga cagaaatcga cccactgtcc acatcacaac attatcttga ctccagcggc gcacagcaca agcggcagaa attetgeece acttgactgt aacaaacat agtcgcacta caccacaccg atctcgctct gcttgcgggc ctacacaata aagatcagag tgccagccat tctattctgg aggcatgctg tctctctctc aaaataaaa tgtacttgtt cageteagag cacacatggt teggtaceag tattgtataa atatttgaac gaagtgatgc gtacgtgttc cagaggtaac tegttgetge gccacgcatt ccaaagcgag tccacaaaga tttcagagca taccgagtaa aatccgttga acttacctgt ggcattctgg aggtgccact taggtgtcat tagggtatgt gacttaaggc ccatgggtct ctctttacca cgatgcccat tgcgagcttc cagacctagg actcaaagac accacccgtc agattgaaga taacatgttg tggctaataa gccttctagt agacaatagc tetetetete tetetetete catcacttta attgatgcct ggtcccatgc ccagacttag gccgaacgag ctaacacagt aaaatttatt gegtteeege gcgagcattc cgaatctcgg gccgtggcgg gcagatggaa tgataagagt tgagcagtac ctgttccttt ttacatgatt aacctdccaa aatcqtcacc tegggcaata cttatggtat ccatggtata gtacaatatg tcatctagtc tgccaaattc actcattgtc ctagcaacta atgetteaac cacttccagg ctactagctc agatcacttc tgaccctgga attgtctgag aggattggga attaattatg ccctatccaa tecgageeet acagtggagg acgttggctt aagcgaatgc ctacgccata gatctgctgt tetetetete tetetetete tgccttttat cttagaaagt gaagttttgc 1621 2341 2461 2641 1441 1561 1681 1741 1801 1861 1921 1981 2041 2101 2161 2221 2281 2401 2521 2581 2701 2761 2821 2881 2941 3001 3061 3121 3181 3241 3301 3361 3421 3481

2

25

30

35

2

aaccaactta tcccctaatg ggctcgtatg ccatgattac ccgcggtggc gcttatcgat tccattcgac tccaaaggag caaaacagct tctttccctc acaattaatg caagttctga cttcatatcc atgagactgt agagccttag gcatttcata ttgtactgct ccttccaact caagattttc atgaaacaaa atatatgttt gctagtygta tcagaatggt tgcactacca ctgaagaatt ccagctgaat ctttgaggaa ttgtatggag ataactgaaa cccataaaa ctctagtctc aaataataqq cacattgtta attgaaacta tggageteca gatccaagag cactgatgaa tgtcatatga tegatateaa ttatgcttcc acagctatga tgaactcaca cacaatccca gcaatggcaa aatggttgct caaacagaga ctcacctgga gtctttattc aaagggcagc acaaaactat tcccaggtaa aacaaaatct agactgacat ctatcatcac aaaaagcagg tttccacagt taaaaacaaa aaatattatt tatgtatgtc tttacagagg ctaaacttca ctattatttc aatacagaac aatagcttct gaacactcct catggaagat ttggtttgca tttgttataa agcattcaca gaaagtgcgt ctgctcttca tttgtcaaga tgtcctgcct agcatggctc agatgcacgt tgttacctac aaccatagta tgttttccat ctctaaggtt cccatatttt gctcaatgga acatgagcaa ttagcagaac aggcagaaaa attcaatcca aaatggacct tgaccctgag tcagcaaatt gctttacact cacacaggaa gaacaaagc aggtggacta ggggaaacaa tctctccatg atacacatct ctgcaggaat caaagaggag aagcttctgc tccagcaaaa accaaggtat gcgacaaaga gacctgigag aggtctccga ctggcatagg acaatgtgtg ttgccaccta atagaagcaa aagacgtaat agaatgtaga aaggagaa acaaaagaaa atggcactga cagtgatctc caaacagtgc aaattgagta tttctcctat ctcgaaccat ttaagttatt aacgccattg tggctacagc tataactttc ggcacccag ataacaattt tcactaaagg atcccccggg aaatgatttc attgtagtgg acagacccac aatgaaatgc tctctctaga taattctcgt gttaaggtgg gaaaaatgcc tcttcaaaca ctcacaatat gtgattggat aagaggettt gcattgtcaa gggcgatcca ggaagcacat ggtgtagaca cttagcaagg acagagattg atttttaaat atgattgtcc tgccaggcca aacacatttg ctataggaaa acttctggtc tgcttcttta tgtaactgaa tcctgatgga tttgctctcc agaaaaaag tggtaaggta gacggaatgt ttgtgagcgg gaactagtgg ccatcactga aaatcattaa tcactcatta caattaaccc tegagetgea tacctgattt gacctgagca aatttccaca aaagtgcttt cttgggacaa tgttgcttac taatgtcaaa gtgctgggca agtggctgaa acatttactg tgatacctgt tcatatcata agtcctcaga attctgctta ttttcccc tttgtcaatt attgtatatt ctctgtcatc ctgtatcctc aatattccag ttctcactta ttcacatgca aatgaaacag catccaacag ttgctaatta attatctggt aaaactaata aagctttcct attttggtaa gtgagttagc ttgtgtggaa cttgacctga gccaagcgcg ggccgctcta atgaaaaggc accgctgacc cagagagaa attcatctgt tgctccttcc gtaggtttta ttttggataa tggtttaggg ttgcacagct gcaagaagat ggtagagata ctgacctttt agactttctt aaaaatcagt tcacaaaagg tcttgtctta taggaaagta ttctttactg ttcacaattc atgttgtact tatatttgct cactgcaagt ctatgttttg agttggctcc taactaataa taaagatccc tettetetee cccagaatta gatagattta tcacgtcttg agaaggatca 3781 3841 3901 3961 4021 4081 4141 4201 4261 4321 4381 444I 4501 4561 4621 4681 4741 4801 1861 4921 4981 5041 5101 5161 5221 5281 5401 5461 5581 5341 5641 5701 5521 5761 5821

accacgggga ccggctgtcc gtggacaaga cctgaggtca ctgcccccat agtggatcaa ttcttcgtgt caggetecag acaacagact aacacgttat accctggtca tcctccaaga cccgaaccgg agcagcttgg gcacctgaac ctcatgatct ccgcgggagg caggactggc cccatcgaga ggcttctatc tacaagacca accgtggaca gctctgcaca ccgaaaagc tttaatatt ggtacccagg acggccagga cagaagtcag atccctgaga gcccaggtgg teggtcacte tgtctcataa caggtcacgc cggagggagg agagteteet agggaggtgt aaggtcaaac cadcaaqctc gggctcagcc aaggatattt agagccgaaa ggggtccctt ctgggtccgc tgatggtggg atattactgt gggccaggga ggactacttc cgtgccctcc ccacgaagac tatatctaaa tgattcaaaa cctggcaccc gcacaccttt caacaccaag accgtgccca caaggacacc caagacaaag cgtcctgcac cctcccagcc ggtgtacacc cctggtcaaa ggagaacaac gatgcatgag agcgccagag cccaggacaa ttggtaccag acceteeggg tatcagtggg tggttatcat ggctgcccc cacactggtg ctacagctgc agatagcagc atgttcaccg tatggtgtca tgaatttctc tgtttaccta cggctgcgcc tctcaagaga cggtcttccc gcctggtcaa acgtggtgac acacatgccc tggacgtgag cctggaaggc taaagccggg cctggatgag acacagccgt ccccgaattg ccagcgjcgt acaagcccag cccaaaacc tgcataatgc tgtcagtgtc ccaccttgac atattcccca aaagcaaaat gegteeteac ccaacaagc gagaaccaca tcttcctcta ctccgggtaa aatatgttta acagcaaacg ctgacagcag ccctgcaga gcctgacctg atgggcagcc catgctccgt gtcagcccaa ccaacaaggc aaagcaacaa cccacaaaa aaatgatttc ggccgtatta aaagccgagg gttatccctc ttccttagca aacgtgaatc gacaaaactc ttcctcttcc tegtgtgtgg atttcgacga ggaggcttgg agattcacca aagggcccat ggcgccctga tgcgtggtgg gggcagccc atctatgagg ccctccaaac aagacagtgg ttcaggctat tcaacagttt ttcagaaacg gccctgggct ggcgtggagg cgtgtggtca tgcaaggtct aaccaggtca gacggctcct aacgtcttct gggacaatgg tgttactcaa accetcctag tgggagagca ctctccctgt ccaccctcgg ttgccagaaa gagetteaag gtgacagtgg cagtggaagt caatcccatt cgtgaaaggc tgtggtggtc tetggetttg ggagtgggtc gaatagcctg atttggggga agcetecace gggcacagcg aggactctac ctacatctgc accgtcagtc tgaggtcaca gtacgtggac cagcacgtac ggagtacaag tgggtcacaa gtggaactca caaatcttgt cgccgtggag gctggactcc gcagcagggg tggagatgca gaccaagetg ctcctctgag cccgggagcc gacgettgag caccgtggag cgtgacttct aatttaatat ggagtcgggg tggattcact caaagccaaa gctgaccaag gcagaagagc gctgacacag ctccagctca tgactactac gaccaccaca aaatctaacc cgtttattaa ttatgataac ccgtctcctc tgcagctgca gtgcagcctc tgacggtgtc tacagtcctc tcctgggggg aacctgtggg tcattcatct tegetttggt ggaagggct atgctgcacc atctgcaaat gcacctctgg gcacccagac aagttgagcc agttcaactg agcagtacaa tgaatggcaa aaaccatctc cccgggatga ccagcgacat cgcctcccgt tcacctgctc gccaggcccc gattetetgg tcagcggagg gtgactccta cgggagtgga acctgagcct atgaagggag cccggaccc agagcaggtg accactacac tttcctatga aagatgaagg tgttcccacc 6241 6301 6541 6601 1999 6721 6781 6841 6961 7021 7081 7141 7501 7561 7681 6361 6421 6481 6901 7201 7261 7321 7381 7441 7621 7741 7801 7861 7921 7981 8041 8101 8161 8281 2 15

೫

25

30

gaagggccct

cttcctcgtc cccttgctg gttgcgcagc acatctaaat cagttcaata gttttgtatt gggcccggta tttacaacgt tacacattta tataaatcaa ccactattaa ttaaattttt ggcccactac ccaaaatcat aaagcttatq tacatgcgaa gagatttt cttctttatc taacatcatt aggggttaac ctgatgactc aatccaaatc tgggccaata ggcaaatgca accagucaat gtggctattc ctggattggc atgaaaagcc gccggaagca tgtataccga tggtcatagc gagttgagat atoggccaac actgactcgc gtaatacggt ctataaagat ttectetgee actecteteg taaaatctca ttccccagca gagagaaatc attaacaggt cctcgagggg tggccgtcgt aaaattcgcg ttccgtcagg gctatttacc tttcgcggaa ttgcagcaca cttcccaaca caaaatccct gaacaagagt tcagggcgat ccttaactta atgattttta catategeaa gaagctaaat ctaaaccaat cctgagcttg caaacagtaa acgctaaaaa aagcaccttg cgateceaee cgcccattta aagacccgta ccacaccgtg aaataagtga caacatacga gcattaatga tteetegete ggcgtaatca cacattaatt ctcaaaggcg taggeteege agcaaaaggc cccgacagga tegegecatg ccattgctgc caqcaaqaa agcacattcc ctagatcctg gcgcgctcac cttaatcgcc accgatcgcc atattttgtt ccgaaatcgg tgctgcccat cattgcctcc ttccagtttg agaattaaca ccaatttatt ggtcattata tgtttactcc aaaccg\*cta gctggttatg ggttttattt cagtaaaacg taaataacag aatccctgta tcagatataa gcaggtaaag gtaaaagacc tgtaatagca gccgttttt tggcatcgtt ttgcgcgctt caattccaca gctcttccgc agaacatgtg ggtggcgaaa tgagctaact cgtgccagct tatcagctca cgtttttcca gggccctctg ggaggaaact gctactgcca agagtcgtta acttggctac tacatatgcc cacattatcc gtcgtattac agaggcccgc gtaagcgtta aaccaatagg cgttacccaa aaagggcgaa tgtgtatcca taggggattc atcagtgctc agggtcaacg ttactcaatg gataataata ttgagtgttg ataaaaagg aaaatagata ttatcaagag ataactaage acttgtctcc tgaatgatta gctcaccaat tccctgtaat taaccaacct gagcagtact tcacgatttc tgagggttaa tatccgctca gcctaatgag aatactgagt aatcaacaaa cgtattgggc gegttgetgg tcaagtcaga ggaaacctgt cggcgagcgg aacgcaggaa ttttgaaggg caaaatggat tegetteace cctatagtga ctccaacgtc tatgacaaga agatttttca ccaggcacca gtaatagcga ctcatttttt cgagataggg gatgtgcaga tggcttctt aaaaccctgg aatggaaatt gcttaaaaac gagettgeeg aaagataaat cctttggg cattggtaag cagggaaac tcgatagcag aaattggtag tagacatcac caataaatcc aaaggettgg ctattcatca ctgaaataat ttccctttag gtgaaattgt tttccagtcg taaaaaggcc aaatcgacgc agcctggggt aggcggtttg cgttcggctg atcaggggat cgtgactggg actctgaatg cccaattcgc acqqaacact gggtttcctg caggccagtg ttttccaaag atctgtgctt gccagctggc ctgaatggcg agaacgtgga gttaaatcag aagaatagac tccgggatca atgatgatgt aaacctaaaa attgagcttg gtaaaaaatg tggtgacgaa cagccatccc acaccggttg atgaaggtca tttgtttgga ttcctgccac aaaattaaaa ctactatctg tcagcttttg aaccatcatg tatcaatgcg tgtttcctgt taaagtgtaa cactgcccgc gcgcgggag tatccacaga ccaggaaccq agcatcacaa gegeteggt 840T 8461 8581 8641 8701 8761 8941 8821 8881 9181 9241 1006 9061 9121 9301 9361 9421 9481 9541 9601 9721 9841 1996 10201 9781 9901 9961 10021 10081 10141 10261 10381 10321 10441 10501

25

	10681	accaggcgtt	teceectgga	ageteeeteg	tgagatataa	tgttccgacc	ctgccgctta	
	10741		gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	
	10801		cagttcggtg	taggtcgttc	gctccaagct	gggctgtgtg	cacgaacccc	
	10861		cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtcc	aacccggtaa	
S	10921	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	
	10981		tacagagttc	ttgaagtggt	ggcctaacta	cggctacact	agaaggacag	
	11041	_	ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	
	11101	٠.	acaaaccacc	gctggtagcg	gtggttttt	tgtttgcaag	cagcagatta	
	11161	_	aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	
10	11221	agtggaacga	aaactcacgt	taagggattt	tggtcatgag	attatcaaaa	aggatcttca	
	11281	_	tttaaattaa	aaatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	
	11341		cagttaccaa	tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	
	11401		catagttgcc	tgactccccg	tcgtgtagat	aactacgata	cgggagggct	
	11461	taccatctgg	cccagtgct	gcaatgatac	cgcgagaccc		gctccagatt	
15	11521	tatcagcaat	aaaccagcca	gccggaaggg	ccgagcgcag	aagtggtcct	gcaactttat	
	11581	cegeetecat	ccagtctatt	aattgttgcc	gggaagctag	agtaagtagt	tegecagtta	
	11641	atagtttgcg	caacgttgtt	gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	
	11701		attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	tcccccatgt	
	11761		ageggttage	teetteggte	ctccgatcgt	tgtcagaagt	aagttggccg	
20	11821	cagtgttatc	actcatggtt	atggcagcac	tgcataattc	tcttactgtc	atgccatccg	
	11881	taagatgett	ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	
	11941	•	ttgatattga	ccggcgtcaa	tacgggataa	taccgcgcca	catagcagaa	
	12001	-	gctcatcatt	ggaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	
1	12061	cgctgttgag	atccagttcg	atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	
25	12121		cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	
	12181	gaataagggc	gacacggaaa	tgttgaatac	tcatactctt	cctttttcaa	tattattgaa	
	12241	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	
	12301	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc	ಇದ		
30 SEQ		:104 (pTnMOI	ID NO:104 (pTnMOD (CMV-prepro-HCPro-CPA))	ro-HCPro-CPA	(7			
	-	ctgacgcgcc	ctgtagcggc	gcattaagcg	cggcgggtgt	ggtggttacg	cgcagcgtga	
	61	_	tgccagcgcc	ctagcgcccg	ctcctttcgc	tttcttccct	tcctttctcg	
	121		cggcatcaga	ttggctattg	gccattgcat		catatcataa	
!	181	tatgtacatt	tatattggct	catgtccaac	attaccgcca		gattattgac	
35	241	tagttattaa	tagtaatcaa	ttacggggtc	attagttcat	agcccatata	tggagttccg	
	301		cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	cccgcccatt	
	361	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	
	421	atgggtggag	tatttacggt	aaactgccca		catcaagtgt	atcatatgcc	
	481	aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	

tegetattae actcacgggg aaaatcaacg cctggagacg tccgcggccg ccgcctatag ggtgtgggtt aatccataac cttcagagac aattcacata gatetecacg agettecaca cttgctccta gccgcacaag gtaggcgtgt cttggggcct cacggctgac tgttgtattc cagtgtagtc gactaacaga attgttaggt caacagctta acttggccgt gcaaaacga agccgacctt acttgatatt ctttalctgt ttatgagaaa ctttaaagtg aagaggaaa acatgatatg tccaatctca ctcgacacgg gccatgggtt tatctattcg cggactaggc cttaggcatg ttttggcacc ctgtttttgg tagcggtttg caaatgggcg cgtcagatcg cgatccagcc gacgtaagta ttagcctata ttccattact atactctgtc attatttaca catagcgtgg taggaagata ccaccagtgt attgggctcg cggtggaggg ccatgtgtga gtagcggcgg gcagctgagt ctcttaccga atagctgaca taaaacgact gaatcgaccg atattogtga ctgttactct aaagctcatg accaatttct aaaatcagcg ggcatgctag tcagcaactt tgaaattcga acacccaaac aacttgttaa gtgatgctgg taagtcgagt ctaaaagcaa cggcaaagga gtcctgccta agaaacaagg tcatgctgct caacagttcg tttagtgaac gtacatctac tgggagtttg tgggcgtgga cccattgacg acaccgggac tgccaagagt atggtatagc ctatatgcca gcatgctata tgacgatact ggtcccattt ttttattaaa ctcttc+ccg atgeceacea tcatgytcgc gagcgtggag gtgctgttaa caccgtcgga aacatcaaac gaagatgcag accagacata gaattacact aaactctca gtataccgtt gactggtctg cacggtcgtt ctcattgtca tggtactggt tttaggctat aagaggctga atctcgctct aaaggccgaa tggaaaccta tactcagcgt cgttttgata gacttgaaaa gttcatgctc aacgtactct acaactccgc ctacttggca agtacatcaa ggattccccg gctataggtg tacaggatgg ttgacgtcaa gcagageteg tccatagaag ccctattgg tctctattgg cggacatggg gcacagcaca tggctcttat ctccagcggc gtctgaaaat agcggcagaa tgcccgcagt tcccgttgcg tegttgetge egegegee tttctgcagt acttgactgt tgtacttgtt tgcgagcttc agtcgcacta atgttcaaag caccacaccg agcggaaac attetgeece aacaaacat gcgactcgct gaagctgggt acctaaaatc aaccttccga cageteagag gcttgcgggc cagaaatcga aaatgtcgta tgggactttc cggttttggc tgttttgacc attggaacgc ctccaccca gtctatataa cacacccctt gcttccttat gccacaacta tctgtattt ggtcccatgc attgaccact ccgtcccccq gtacgtgttc ccagacttag gacttaaggc cagaggtaac ctctttacca tagggtatgt ccaaagcgag tccacaaaga ccatgggtct gccacgcatt tttcagagca taccgagtaa aatccgttga actcaaagac tattgtataa accacccatc cgatgcccat cagacctagg aagcgaatgc agattgaaga acttacctgt gccgaacgag taacatgttg ctaacacagt actctatagg catgacetta ccatccacgc atttccaagt ggactttcca acggtgggag ggaacggtgc tgacacggac catggtgatg atacacccc attgaccatt atggctctt tacaacaacg cgaatctcgg gccgtggcgg acagtggagg gcagatggaa tgagcagtac ctgttccttt tccgagcct ttacatgatt tgataagagt aatcgtcacc acgttggctt aacctgccaa tcgggcaata cttatggtat gegtteeege gcgagcattc ccatggtata gtacaatatg tgccaaattc actcattgtc ctagcaacta atgetteaac tcatctagtc ctacgccata cacttccagg 1021 901 1081 1141 1201 1261 1441 1501 1561 1321 1381 1621 1861 1681 1741 1801 1921 1981 2041 2281 2101 2161 2221 2341 2401 2461 2521 2581

ggtacctctc attagccttg ggctcgtatg gcttatcgat ataatgatcc gtttttgtg gtgccttcct attgcatcgc tetetetete tgaccaaagg ataaqcaqca gttggtctgc aaaaccttat aaccaactta tecectaatg tcccggtaat ccatgattac ccacaataac tccatatcat ttgattattg tatggagttc cccccdccca gtatcatatg ttatgcccag ccaaaatcaa ctaatccata tccttcagag agcaaggggg ccattgacgt catogotatt tgactcacgg cggtaggcgt cgcctggaga cctccqcqqc gacttggggc taggtgtggg caaattcaca taccgcctat aaatgaggaa ggggcagcac gggctctatg attgacccgg agtgcctgtt cttgaaaaa tgtcatatga tggagctcca tegatateaa catgttgaca atagcccata tagggacttt tgttttggca acttactcqt ggaaattatg ctgtgtgttg cetetetete ttaacaaatg gataataata gatccaagag cactgatgaa ttatgcttcc acagctatga atacgttgta cgcccaacga tacatcaagt ccgcctggca acgtattagt gatagcggtt cgcaaatggg accgatccag gcttagccta cttccatta caatactctg cactacaca tactqttttt ttattattta accgtcagat gtgacgtaag acaattactc acaaqqqaaq ctgttgtttg tctctcggta gtgctgaaga ttggaatgaa ctgcaggaat ggtgacgata tacgetttgg ctctagagat tttcctaata 9999t9999t gggatgcggt aaaaactgat ttatattatt tcagcaaatt tgaccutgag atacacatct gctttacact cacacaggaa gaacaaagc tggccattgc acattaccgc tcattagttc cctggctgac gtaacgccaa cacttggcag ggtaaatggc cagtacatct aatgggcgtg aatgggagtt gccccattga cgtttagtga agacaccggg cqtqccaaga atgcatgcta tgatggtata ggctatatgc ggggtcccat cccactgtcc tctattctgg gttaaggtgg ctacacaata aagatcagag aaaataaaa acatcacaac attatcttga aacgccattg tcactaaagg gattggctat ctcatgtcca gtaaactgcc cattgacgtc cacacatggt tgccagccat aggcatgctg tetetetete teggtaceag ctatcattgg taattctcgt ggcacccag atcccccggg aattacgggg tgttcccata cgtcaatgac taacaactcc cctccataga geggatteee tttggctctt tttacaggat ataacaattt aaatggcccg tcctacttgg gcagtacatc aagcagagct atqctatagg ctcccctatt tatctctatt taggtgtcat agacaatagc gacggaatgt gaactagtgg tggctaataa aggtgccact tetetetete tetetetete atatttgaac gaagtgatgc tggtaaggta aaatcattaa ttgtgagcgg taatgacgta tatgggactt tgcggttttg aggtctatat gctgttttga ggcacaccc ccgcttcctt ttgccacaac actctgtatt ggcattctgg gccttctagt catcacttta tcactcatta caattaaccc tegageatea tttatattgg aatagtaatc aacttacggt agtatttacg ccctattga caaaatgtcg gcattggaac ttattgacca aaaatttatt attgatgcct gtcttcaccc gaagttttgc ctactagete agatcacttc gatctgctgt tgaccctgga attgtctgag aggattggga tetetetete tetetetete tgccttttat attaattatg cttagaaagt ccctatccaa aatacattac aagctttcct attttggtaa gtgagttagc ttgtgtggaa gccaagcgcg ggccgctcta aatatgtaca actagttatt cgcgttacat ttgacgtcaa caatgggtgg accatggtga ggatttccaa cgggactttc cgccatccac acatggctct actgacacgg accgctgacc tcaagtacgc tacatgacct gtacggtggg cgggaacggt agactctata ctatacaccc ttattgacca 3181 3241 3301 3421 3541 3601 3841 3901 4141 3001 3061 3121 3361 3481 3661 3721 3781 3961 4021 4081 4201 4261 4321 4381 4441 4501 4561 4621 4681 4741 4861 4921 4981 5041 5101 5161 4801

೫

25

30

35

15

tatacaacaa

ggagcttcca tecttgetee gtgccgcaca cgcacggctg gttgttgtat ggcagtgtag cagactaaca tcatctcgtq tttggttctg gctgcaggag agcctctgga ggggctggag tgcacccgtg ctctgggggc tattaaaatt gcaaatgaat gataacattt ctcctcagcc gtcctcagga ggtgtcgtgg ccagacctac tgagcccaaa 633633acc3 gacccutgag gtacaacagc tggcaaggag cgacatcgcc caactggtac catctccaaa ggatgagetg tecegtgetg caggtggcag ctacacgcag tgagctgaca cctgtggtg aggtgactac aacatagcgt agattgggct tegtgttege cggtagcggc gctcggcagc caccaccagt aggcagctga aacggtggag taatagctga gatcaatcat atatttcqtt cccaggtgca tatactgtga ctccagggaa cagactatgc ccaagagcac cgttatatct tggtcaccgt aaccggtgac cggggattat ctgtcctaca gcttgggcac tgatctcccg acaagaaagt ctgaactcct aggtcaagtt gggaggagca tcgagaaac cccatcccg actggctgaa tggacaagag ggatcacctg agagattete tctatcccag agaccacgcc tgcacaacca agctttccta caggccaggc tggaagatga gtttttatta ggctcttctc gctcatggtc caatgcccac atgagcgtgg aagaagatgc ccaccagaca gtcaccgtcg tecettagag cggtgctgtt tctaaattta atattttct ccgaaaggta gtccgchagg ggtggyacaa tcaaaaaca tactgtacca cagggaaccc gcaccetect tacttccccg acctttccgg accaaggtgg tgcccagcac tacaccctgc ccctccagca gacaccetea gaagaccctg ctgcaccagg acaaagccgc ccagccccca gtcaaaggct aacaactaca aagctcaccg catgaggete ccagagccga caaacggcca cagcagaagt gggatccctg ggggcccagg cgtgcccgca tccggacatg aggcacagca gtgtctgaaa gcagcggcag actcccgttg tacctatata tttctcaagg cttccccctg cggcgtgcac gcctccagcg 6262626226 cttttctgca tgcgccagag gatgagctgg caaaattgat gaattggggc ggtcaaggac aagagatgat agccgtatat ggtgaccgtg gcccagcaac aaactcacac atgcccaccg aaaacccaag cgtgagccac cctctacage ttattggtac taatgccaag cctcaccgtc caaagccctc accacaggtg gacctgcctg gcagccggag gtccccagga ctccgtgatg acgaccetee gggtaaagcg gactatcagt aagacttaag ttccatgggt cagtttcggc cgccgtcccc gggtacgtgt ctggtcccat ggccagactt ggtagggtat gtcagaggta actcgttgct gtgtggtgtt cgacgatgaa ccgaggacac gcttggtaaa tcaccatctc tecetecee gaaacgcctg gcccatcggt ccctgaccag gtattaaaag tgggctgcct ttagcaacgt tgaatcacaa tattacacca tggtggtgga tggaggtgca tggtcagcgt tetteteatg aggtctccaa agccccgaga aggtcagcct agagcaatgg ccetgtetee geteettett cggtgtcagt aaaaatatgt aggacagcaa tggccacctt taacagtgga cgcgaatctc acgcagatgg tctgataaga catccgagcc aggccgtggc tctgagcagt gactgttcct acttcttcgt taatattt gctttgtcaa ttcactttca tgggtcggcc aaaggcagat gggggagtta tccaccaagg tcggggggag agcctgaaag aactcaggcg acagcggccc ctctacttcc atctgcaacg tcttgtgaca tcagtcttcc gtcacatgcg gtggacggcg acgtaccgtg tacaagtgca gccaaagggc accaagaacc gtggagtggg gcattgccag tcagggacaa gactccgacg cadccacct gtcatctatg caggggaacg aagagcctct 5341 5401 5461 5521 5641 5761 5821 5881 5941 6001 1909 6181 6121 6241 6301 6541 6361 6421 6481 6601 1999 6721 6781 6901 6841 6961 7081 7021 7141 7201 7261 7321 7381

25

30

35

15

ggagaccacc atgtggcttc gtgccaggca accetectet gagcaccgtg actgatgtgc tgactgggaa gaatggcgaa ggggattcat cctttttqaa ctgagattt aagcaaaatg cttcgcttc caattcgccc cagctggcgt cgggatcata cctgacgctt taaatcagct gaatagaccg aacgtggact gatgatgtgc tgagcitgaa gaaggtcatc acctaaaaga aaaaaatgcc gtgacgaaat gccatcccaa accggttgca tgtttggata aattaaaaca actatctgca ccatcatgct cctqccacaa taagtgactc aggcgggagt cgcatgaagg ccgcggaggg agggaagggc atattttcca attatctgtg ccctttcgc ctctqttccc gctacctgag cgcacggaac aatgggtttc ctgcaggcca gcccggtacc tacaacgtcg gtcactctga tgcgcagcct aaatttttgt taaatcaaaa actattaaag ccgtcaggaa agcttatgat ccactactc aaaatcatta catgcgaaaa ggctttttat tctttatcgt gggttaacat ggccaataac gatgactett cagccaataa tataccgatc acatcatttg tccaaatcca caaatgcact ggctattctt gaaagccaa ggattggcta cggaagcata ccctcggtca gtgtgtctca agccccgtca geggeeagea tgccaggtca tegtgeece gecetteete tcgagggggg gcagcacatc aattcgcgtt aaatccctta tatttaccgc tcaacatcta gcaccccttg atccagttca ggtgttttgt gccgtcgttt acaagagtcc agggcgatgg gatttttacc tatcgcaata tcccaacagt agctaaatct togoggaata tgagcttgag aaaccaataa aacagtaatg gcaccttgct atcccaccac gctaaaaagg ataagtgatg cgtaatcatg cccatttagt gacccgtaat acaccgtgct acatacgage ggcagatagc caacaagtac aagctacagc caaggetgee agaatgttca atgactcctc ggccacactg tgcttcctct teceetegae attttgttaa ccagtttgga aatttattgc aaataaatc teetteecea catgagagaa ctgattaaca gcgctcactg taatcgcctt gaaatcggca accetctatc ttaacttaat cgatcgccct tggttatgca ttttatttga aattaacatt aataacagca gcatcgttaa tcattatatt tttactccc gtaaaacgct aggtaaagcg taatagcacc tccctgtaaa agatataaac cgtttttcq aaaagaccaa gegegettgg attccacaca aacaaagcaa taggtcagcc tggcccctgc tggcctggaa agtcccacaa acttcgcgcc cgtattacgc aagccaacaa ctgccattgc ccacagcaag tactgctgcc gccctagatc teceattgee aggcccgcac aagcgttaat ccaataggcc gagtgttgtt tgtatccacc ttaagcacat ttacccaact agggcgaaaa ggtcaacgag aaaaaaggcc actcaatggc aatagatagg aatgattata cctgtaatgc atcaagaggg ttgtctcctg taataataca tcaccaataa accaaccttc gcagtactgc tcaacaaatg tactgagtgt acgatttctg agggttaatt tccgctcaca ctgaccgtcc gaggagette gagcagtgga gccgtgacag acaccctcca gggggaggaa agagggcct tttgctactg gagaagacag tcaagagtcg gattacatat aaccctggcg ccaacttggc acceacatta tatagtgagt aatagcgaag tggaaattgt cattttttaa agatagggtt tgacaagatg ccaacgtcaa cagtgctcag gcttgccgat agataaataa ctcttgggtt aactaagcac gatagcagga ttggtaaggc aggcttggaa ttaaaaactt attggtagtg gacatcactc gggaaaacta ataaatccga attcatcatc gaaataataa cctttagtg gaaattgtta 7681 7741 7801 7861 7921 8041 8101 7981 8161 8341 8521 8221 8281 8401 8581 8641 8701 8761 8821 8461 8881 8941 9001 9061 9121 9181 9241 9361 9301 9421 9661 9481 1096

15

2

catagggagg

cagtggttat

caactgacag

tactgttact

20

25

ggatacctgt aggtatctca ggcggtgcta ctgcccgctt gcggggagag cgctcggtcg tccacagaat aggaaccgta catcacaaaa caggogttto cacgacttat tttggtatct gttcag*c*ccg cgcagaaaa tagatccttt cgttcatcca tccggcaaac tggaacgaaa tggtctgaca ccatctggcc tcagcaataa agtttgcgca atggcttcat gtgttatcac gcctccatcc tqcaaraaaq agatgctttt ctgttgagat cgaccgagtt ttaaaagtgc actttcacca ataagggcga atttatcagg caaatagggg gttgcgctca aatacggtta gcaaaaggcc cggccaacgc ccctgacgag gccgcttacc tgactcgctg ataaagatac ctcacgctgt cccggtaaga gaggtatgta aaggacagta tagctcttga cgaaccccc gcagattacg tgacgctcag gatetteace tgagtaaact ctgtctattt ggagggetta tccagattta gtggtcctgc aactttatcc gccagttaat cccatgttg gtcgtttggt gttggccgca gatcttaccq aaaaaaggaa gccatccgta gtgtatgcgg taggagagg agcatcttt gaaaaataaa cattaattgc cctcqctcac caaaggcggt caaaaggcca ggctccgccc cgacaggact tttctcatag gctgtgtgca ttagcagagc ttgagtccaa gctacactag tttgcaagca ctacggggtc attaatqaat ttccgacct tctcagcgat aaagagttgg tatcaaaaag aaagtatata ctacgatacg geteaeegge taagtagttc tgtcacgctc ttacatgatc tcagaagtaa ttactgtcat tctgagaata ccgcgccaca aactctcaag aaaatgccgc tttttcaata aatgtattta actgatcttc agctaactca aacatgtgag tgccagctgc tetteegett tttttccata tggcgaaacc agatatactg agcgtggcgc aactatogto ggtaacagga cctaactacg ggttttttg gtcatgagat gtgtagataa tcagctcact tccaagctgg accttcggaa ggcatcgtgg tcaaggcgag ttgatcttt atgaagtttt aaatcaatct gaggcaccta cgagacccac gagcgcagaa gaagctagag ccgatcgttg cataattctc cgggataata tcggggcgaa acaggaaggc tacatatttg cgtgcaccca atactettee accaagtcat aaagtgccac tattgggcgc gcgagcggta cgcaggaag tggtagcggt ctaatgagtg aaacctgtcg gttgctggcg aagtcagagg ctccctcgtg cccttcggga ggtcgttcgc cttatccggt agggattttg cttaatcagt agcagccact gaagtggtgg gaagccagtt agaagatcct acteceegte cggaaggcc cattgctaca cttcggtcct ggcgtcaata catgagcgga aatgataccg ttgttgccgg ttcccaacga ggcagcactg gtgagcaaaa ttgaatactc atttccccga tgagtactca aaaacgttct gtaacccact cctggggtgc tccagtcggg geggtttgeg accgctgcgc cgccactggc aaaccaccgc aaaaggccgc atcgacgete cagagttctt ttcggctgcg caggggataa ccctggaag gttcggtgta gegetetget aaggatctca actcacgtta ccacctttct taaattaaaa gttaccaatg tagttgcctg ccagtgctgc accagccagc agtctattaa acgttgttgc tcagctccgg tcatggttat ctgtgactgg cggttagctc gcgtttctgg ttccgcgcac getettaece tcatcattgg ccagttcgat gttattgtct cacggaaatg 1966 0021 10081 10141 10201 10261 10321 10381 10441 0501 10561 10741 10801 10861 10621 .0681 11281 10921 11041 10981 11101 11161 11221 11341 11401 11461 11521 11641 11701 11581 11761 11941 11821 10 15

20

25

30

35

## **CLAIMS**

We Claim:

5

15

20

25

30

- 1. An isolated polynucleotide comprising at least one gene of interest and one or more pro nucleotide sequences, wherein the at least one gene of interest is operably-linked to a pro nucleotide sequence and each of the at least one gene of interest may be the same or different when more than one one gene of interest is present, wherein the at least one gene of interest encodes a protein useful in a vaccine.
- 2. The polynucleotide of claim 1, wherein a most 5' pro nucleotide sequence of the one or more pro nucleotide sequences is a part of a prepro nucleotide sequence.
  - 3. The polynucleotide of claim 1, wherein two genes of interest and two pro nucleotide sequences are arranged in the following order: a prepro nucleotide sequence, a first gene of interest, a pro nucleotide sequence, and a second gene of interest.
    - 4. The polynucleotide of claim 3, wherein the prepro nucleotide sequence is a cecropin prepro nucleotide sequence and the pro nucleotide sequence is a cecropin pro sequence.
      - 5. The polynucleotide of claim 3, wherein the prepro nucleotide sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4 and the pro nucleotide sequence comprises a sequence shown in SEQ ID NO:1 or SEQ ID NO:2.
      - 6. The polynucleotide of claim 1, wherein a first gene of interest encodes for an antibody heavy chain and a second gene of interest encodes for an antibody light chain.
      - 7. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising two or

5

30

more genes of interest, wherein each gene of interest encodes a part of the multimeric protein, each gene of interest is operably-linked to a pro nucleotide sequence, and each of the two or more genes of interest may be the same or different, wherein each of the two or more genes of interest encodes a protein useful in a vaccine.

- 8. The method of claim 7, wherein a most 5' pro nucleotide sequence of the two or more pro sequences is a part of a prepro nucleotide sequence.
- 9. The method of claim 8, wherein the polynucleotide comprises two genes of interest and two pro nucleotide sequences arranged in the following order: a prepro nucleotide sequence, a first gene of interest, a pro nucleotide sequence, and a second gene of interest.
- 15 10. The method of claim 7, wherein a first gene of interest encodes for an antibody heavy chain and a second gene of interest encodes for an antibody light chain.
- 11. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising a cecropin prepro nucleotide sequence operably-linked to two or more genes of interest, each gene of interest encoding a part of the multimeric protein, wherein the multimeric protein is useful in a vaccine.
- 25 12. The method of claim 11, wherein a first gene of interest is an antibody heavy chain and a second gene of interest is an antibody light chain.
  - 13. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising two or more genes of interest, wherein each gene of interest encodes a part of the multimeric protein and wherein each gene of interest is operably linked to a gene encoding for a cleavage site, wherein the multimeric protein is useful in a vaccine.

PCT/US2003/041269 WO 2004/067743

> The method of claim 13, wherein a transposon-based vector comprises 14. the polynucleotide and further comprises a transposase gene operably linked to a first promoter and wherein;

5

- the first promoter comprises a modified Kozak sequence comprising ACCATG;
- the two or more genes of interest are each operably-linked to one or more additional promoters; and,

10

the two or more genes of interest and their operably-linked c) promoters are flanked by transposase insertion sequences recognized by a transposase encoded by the transposase gene.

15

The method of claim 13, wherein a transposon-based vector comprises 15. the polynucleotide and further comprises a transposase gene operably linked to a first promoter and an avian optimized polyA sequence, and wherein;

the two or more genes of interest are each operably-linked to one or more additional promoters; and,

the two or more genes of interest and their operably-linked b) promoters are flanked by transposase insertion sequences recognized by a transposase encoded by the transposase gene.

20

An animal comprising the isolated polynucleotide of Claim 1. 16.

25

The animal of Claim 16, wherein the animal is a bird. 17.

An egg produced by the animal of Claim 16. 18.

The egg of Claim 18, wherein the egg comprises a multimeric protein 19. encoded by the isolated polynucleotide.

30

The animal of Claim 16, wherein the animal is a mammal. 20.

21.

Milk produced by the mammal of Claim 20.

22. The milk of Claim 21, wherein the milk comprises a multimeric protein encoded by the isolated polynucleotide.

5

10

15

- 23. A method of producing a multimeric protein comprising:
- a) administering to an egg-laying animal a composition comprising the polynucleotide of Claim 1; and,
- b) permitting the one or more genes of interest to be expressed into the multimeric protein, wherein the multimeric protein is useful in a vaccine.
- 24. The method of Claim 23, further comprising
  - a) collecting an egg from the egg-laying animal;
  - b) harvesting egg white comprising the multimeric protein; and,
  - c) purifying the multimeric protein.
- 25. The method of Claim 23, wherein the egg-laying animal is a bird.
- 26. A method of producing a multimeric protein comprising:

20

- a) administering to an intramammary duct system of a mammal a composition comprising the polynucleotide of Claim 1, and,
- b) permitting the one or more genes of interest to be expressed into the multimeric protein, wherein the multimeric protein is useful in a vaccine.

25

- 27. The method of Claim 26, further comprising
- a) collecting milk from the mammal, wherein the milk comprises the multimeric protein; and
  - b) purifying the multimeric protein.

30

28. A method of vaccinating an individual against a disease comprising administering to the individual a composition comprising a protein, a multimeric protein or an antibody or a combination thereof, wherein the

composition is produced in the individual or in another individual made transgenic using transposon-based vectors.

29. Use of a composition comprising a protein, a multimeric protein or an antibody or a combination thereof, wherein the composition is produced in an individual made transgenic using transposon-based vectors, in the preparation of a medicament for use in a vaccine.

5

15

- 30. The polynucleotide of any of the preceding claims, wherein the prepro nucleotide sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.
  - 31. The polynucleotide of any of the preceding claims, wherein the two or more pro nucleotide sequences each comprise a sequence shown in SEQ ID NO:1 or SEQ ID NO:2.
  - 32. The polynucleotide of any of the preceding claims, wherein the prepro nucleotide sequence is a cecropin prepro nucleotide sequence.
- 20 33. The polynucleotide of any of the preceding claims, wherein the pro nucleotide sequence is a cecropin pro nucleotide sequence.

IGURE

Prom prepro Gene of Interest pro Gene of interest

polyA

1/5

FIGURE 2

Γ	SI
	polyA
	Light chain
	pro
	Heavy chain
	prepro
	Oval Prom
	SI

IGURE ?

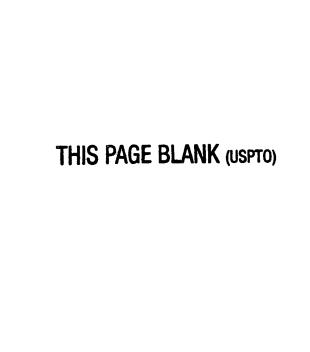
Prom Cecropin prepro Gene of Interest CS Gene of Interest polyA

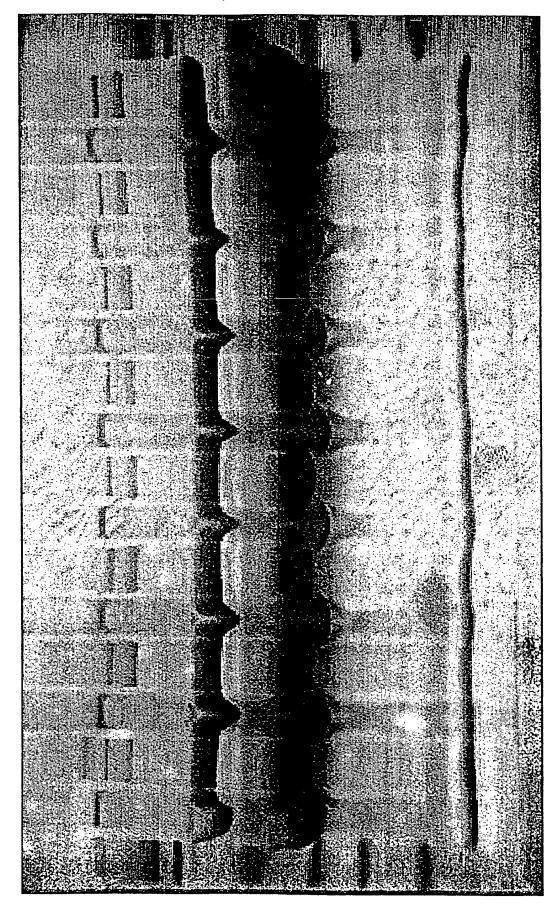
3/5

4/5

GURE 4

IS Prom SS Gene of Interest CS Gene of Interest polyA IS





FIG

## This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS
☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
FADED TEXT OR DRAWING
BLURRED OR ILLEGIBLE TEXT OR DRAWING
☐ SKEWED/SLANTED IMAGES
☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
☐ GRAY SCALE DOCUMENTS
☐ LINES OR MARKS ON ORIGINAL DOCUMENT
☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
□ OTHER:

## IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.